

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 29, 2004, 08:19:58 ; Search time 31.6873 Seconds
(without alignments)
2596.165 Million cell updates/sec

Title: US-09-936-333-27

Perfect score: 4683
Sequence: 1 MGSDRRARKGGGGRDAGL.....PGVYTRLPFRDRIKENTGV 855

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : PIR-79.*
1: PIR1.*
2: PIR2.*
3: PIR3.*
4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3866	82.6	855	2 JC7731	membrane-bound arg
2	724	15.5	1034	1 A53663	enteropeptidase (E
3	709.5	15.2	1035	1 A43090	enteropeptidase (E
4	679.5	14.5	1019	1 A56318	enteropeptidase (E
5	661.5	14.1	1113	2 JE0315	low-density lipopr
6	576.5	12.3	1524	2 T30337	polyprotein - Altr
7	570	12.2	699	1 I54763	Ra-reactive factor
8	528.5	11.3	638	1 KQMSPL	plasma kallikrein
9	507.5	10.8	638	1 KORTPL	plasma kallikrein
10	506.5	10.8	790	1 PLRG	plasma (EC 3.4.21
11	504	10.8	613	2 S15468	complement C3b/C4b
12	495.5	10.6	786	1 A17547	serine proteinase
13	495	10.6	638	1 KQHUP	plasma kallikrein
14	493	10.4	460	2 B61545	plasma (EC 3.4.21
15	488.5	10.4	583	2 A29154	complement factor
16	488	10.4	810	1 PLHU	plasma (EC 3.4.21
17	486	10.4	416	1 KRBH	coagulation factor
18	486	10.4	417	1 S00845	hepsin (EC 3.4.21.
19	485.5	10.4	812	1 PLBO	plasma (EC 3.4.21
20	483.5	10.3	343	1 A57014	proctasin (EC 3.4.
21	483.5	10.3	812	1 PLMS	plasma (EC 3.4.21
22	479.5	10.2	416	1 S33777	hepsin (EC 3.4.21.
23	477	10.2	810	2 B30848	plasma (EC 3.4.21
24	475	10.1	686	1 A59271	Ra-reactive factor
25	474.5	10.1	625	1 KFHU1	coagulation factor
26	471	10.0	275	2 S40007	trypsin (EC 3.4.21
27	469	10.0	810	2 I46260	plasma (EC 3.4.21
28	468	10.0	2616	2 A57096	model protein prec
29	467.5	10.0	2616	2 A57096	model protein prec

30	466.5	10.0	452	1 A30351	coagulation factor
31	463	9.9	1420	2 A32869	apolipoprotein(a)
32	461	9.8	1455	2 A61545	plasma (EC 3.4.21
33	460	9.8	275	2 S40005	trypsin (EC 3.4.21
34	458	9.8	562	1 UKHUT	c-plasminogen acti
35	455	9.7	274	2 S35339	trypsin (EC 3.4.21
36	454.5	9.7	267	2 S40006	trypsin (EC 3.4.21
37	453.5	9.7	560	1 JC4795	plasma hyaluronan-
38	444	9.5	228	1 TRW5X	trypsin-like prote
39	441.5	9.4	1004	2 T30338	oviductin (EC 3.4.
40	440.5	9.4	4548	2 S00657	apoprotein(a) (EC
41	439	9.4	254	1 TRW3J	trypsin-like prote
42	434	9.3	263	2 A21195	chymotrypsin (EC 3
43	432	9.2	242	2 S49489	trypsin (EC 3.4.21
44	432	9.2	263	1 KYRIB	chymotrypsin (EC 3
45	430.5	9.2	247	1 TRDG	trypsin (EC 3.4.21
46	430	9.2	270	2 S56160	mast cell tryptase
47	429.5	9.2	459	2 JQ0419	coagulation factor
48	426	9.1	277	2 S35340	trypsin (EC 3.4.21
49	425.5	9.1	407	1 KFB07	coagulation factor
50	425	9.1	265	2 T15451	hypothetical prote
51	424.5	9.1	231	1 TRPGR	trypsin (EC 3.4.21
52	423.5	9.0	558	2 JC5878	plasma hyaluronan-
53	422	9.0	231	2 S31778	trypsin (EC 3.4.21
54	421.5	9.0	247	1 B25852	trypsin (EC 3.4.21
55	420.5	9.0	247	2 A27547	trypsin (EC 3.4.21
56	420.5	9.0	304	2 S33496	trypsin (EC 3.4.21
57	420	9.0	263	2 A31299	chymotrypsin (EC 3
58	419.5	9.0	705	2 A47246	trypsin (EC 3.4.2
59	419.5	9.0	273	2 C1HURB	complement subcomp
60	419	8.9	246	2 JQ1472	trypsin (EC 3.4.21
61	418	8.9	245	1 KYBOB	chymotrypsin (EC 3
62	417.5	8.9	761	2 JC5759	brain-specific ser
63	417	8.9	246	2 JQ1471	trypsin (EC 3.4.21
64	417	8.9	276	2 A38654	mast cell proteina
65	415.5	8.9	246	1 TRRT1	trypsin (EC 3.4.21
66	415.5	8.9	266	2 S54146	trypsin (EC 3.4.21
67	415.5	8.9	482	1 EXRT	coagulation factor
68	414	8.8	229	1 TRDS	trypsin (EC 3.4.21
69	414	8.8	246	2 B25528	trypsin (EC 3.4.21
70	413.5	8.8	456	1 KXBO	thrombin (EC 3.4.2
71	413.5	8.8	622	1 TBHU	trypsin (EC 3.4.21
72	413	8.8	242	2 S31776	trypsin (EC 3.4.21
73	413	8.8	242	2 S31775	trypsin (EC 3.4.21
74	413	8.8	274	2 JC4171	trypsin (EC 3.4.21
75	412	8.8	247	2 S05494	trypsin (EC 3.4.21
76	412	8.8	275	2 A32410	chymotrypsin (EC 3
77	409	8.7	263	2 S47537	u-plasminogen acti
78	409	8.7	432	1 S18932	trypsin (EC 3.4.21
79	408.5	8.7	247	2 S13813	trypsin (EC 3.4.21
80	408.5	8.7	431	2 S47538	acrosin (EC 3.4.21
81	408	8.7	259	2 I38363	trypsin (EC 3.4.21
82	408	8.7	275	2 B35863	trypsin (EC 3.4.21
83	408	8.7	559	2 A35029	c-plasminogen acti
84	408	8.7	618	2 A35827	thrombin (EC 3.4.2
85	407.5	8.7	264	2 I38336	chymotrypsin-like
86	407	8.7	266	2 JC4850	trypsin-like prote
87	407	8.7	275	2 A35863	trypsin (EC 3.4.21
88	407	8.7	559	2 A29941	c-plasminogen acti
89	406	8.7	245	1 KYBOA	chymotrypsin (EC 3
90	406	8.7	466	1 KFHU7	coagulation factor
91	405.5	8.7	238	2 S31779	trypsin (EC 3.4.21
92	404.5	8.6	237	1 TRCY1	trypsin (EC 3.4.21
93	404.5	8.6	274	2 I47078	coagulation factor
94	404	8.6	247	1 A25852	trypsin (EC 3.4.21
95	403.5	8.6	250	2 S55493	serine proteinase
96	403	8.6	271	1 ELRT2	pancreatic elastase
97	403	8.6	275	2 C35863	trypsin (EC 3.4.21
98	402.5	8.6	246	1 TRRT2	trypsin (EC 3.4.21
99	402.5	8.6	247	2 S12764	trypsin (EC 3.4.21
100	402.5	8.6	461	1 S18994	protein C (activat

ALIGNMENTS

RESULT 1

membrane-bound arginine-specific serine proteinase precursor - rat

C/Species: Rattus norvegicus (Norway rat)

C/Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004

C/Accession: J07731; J07775

R/Kishi, K.; Yamazaki, K.; Yasuda, I.; Yahagi, N.; Ichinose, M.; Tsuchiya, Y.; Achauda, J. Biochem. 130, 425-430, 2001

A/Title: Characterization of a membrane-bound arginine-specific serine protease from rat

A/Reference number: J07731; PMID:21421307; PMID:11530019

A/Accession: J07731

A/Molecule type: mRNA

A/Residues: 1-855 <R1S>

A/Cross-references: UNIPROT:Q9J17; DDBJ:AB049189

A/Experimental source: strain Male, 7-week-old

R/Satoml, S.; Yamazaki, Y.; Tsubuki, S.; Hicomi, Y.; Iwanaga, T.; Fushiki, T.

Biochem. Biophys. Res. Commun. 287, 995-1002, 2001

A/Title: A role for membrane-type serine protease (MT-Sp1) in intestinal epithelial turn

A/Reference number: J07775; PMID:11575963

A/Contents: Small intestine

A/Accession: J07775

A/Molecule type: mRNA

A/Residues: 1-855 <SAT>

A/Cross-references: DDBJ:AB037898

C/Comment: This enzyme, an epithelial-derived, type II integral membrane serine protease of specific proteins or peptides on the brushborder membranes. It also participates in

lial migration and/or cell loss.

C/Genetics:

A:Gene: mt-ep1

A:Map position: basolateral cell surface

C:Superfamily: membrane-bound arginine-specific serine proteinase

C/Keywords: protein digestion

Query Match 82.6%; Score 3866; DB 2; Length 855;
Best Local Similarity 80.8%; Pred. No. 6.1e-245;
Matches 691; Conservative 79; Mismatches 85; Indels 0; Gaps 0;

```

QY 1 MGSRAKGGGPGPDGAGLKYNRHRYKNGLEGVFLPNNVKKYKFKGKVVYLA 60
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MGNRRKAGGGSDPFAGLKYNRLENMNGFEGVEFLPNNAKQYKGGPRVVMVA 60
QY 61 VLIGLLVLLGIGLVNHLQYRDVVOYKVGKVRITNENFVADYKSNSTEFYLSKV 120
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 VVRSFLLSLMAGLLVNHFRYRVRIOKVNGHRIKRTNENFLDYENSTEFYLSKV 120
QY 121 KDALKLLYGVFLGPHYKKSAAVTAFFSEGSVIAVYMSFESIPOHLVEAEERVAE 180
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 KEALKLVSEVPVLGPHYKKSSTVAFSEGSVIAVYMSFESIPHLREVDRAVAVRVT 180
QY 181 LPPARSLKFPVYVSVAFPPDSKTVORTDNGSCFGLHARGVLAFFTPGPFDSPPA 240
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 LPPARSLKFSVLTSVAFPIDEPMLDRTDNGSCFALHARGVTVFTTPGPFNSPYPA 240
QY 241 HARCQMLRGDADSVLSLTFERSFDIACDERGSDLVVYVYTLSPMEHALVOLCGT 300
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 241 HARCQMLRGDADSVLSLTFERSFDVACDGDSDLVVYVLSLMEPHAVVRLCGTSPS 300
QY 301 YNLTFFSSQVLLTLITNTERRHGFGEATFFQIPRMSCGGLRKAQGTFFNSPYRG 360
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 301 YNLTFFSSQVFLVTLITNTERRHGFGEATFFQIPRMSCGGLLSAQAQGTFFNSPY 360
QY 361 PPNIDCWNIIEVNNQVYKRFKFFYLREGVNPGTCKPVYENKRYKCGERSQFVY 420
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 361 PPNIDCWNIIEVNNQVYKRFKFFYLREGVNPGTCKPVYENKRYKCGERSQFVY 420
QY 421 NSNKITVRFHSDQSTDTGFLAETLYSDSDPCPGQFTCTRGTCIRKELRCDGADCT 480
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 421 NSNKITVRFHSDQSTDTGFLAETLYSDSDPCPGQFTCTRGTCIRKELRCDGADCP 480
QY 481 SDEINSCDAGHQTCKNKKCPFLFWVCDSVNDGSDSGSCSPAQTFRCSNGKCLSK 540
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

RESULT 2

enteropeptidase (EC 3.4.21.9) precursor [validated] - pig

N/Alternate names: enterokinase

C/Species: Sus scrofa domestica (domestic pig)

C/Date: 07-Oct-1994 #sequence_revision 09-Aug-1996 #text_change 28-Apr-2003

C/Accession: A53663

R/Matsushima, W.; Ichinose, M.; Yahagi, N.; Kakei, N.; Tsukada, S.; Miki, K.; Kurokawa, I.

J. Biol. Chem. 269, 19976-19982, 1994

A/Title: Structural characterization of porcine enteropeptidase.

A/Reference number: A53663; PMID:94327548; PMID:8051081

A/Molecule type: mRNA

A/Residues: 1-1034 <MAT>

A/Cross-references: GB:030799; NID:G505122; PIDN:BA06459.1; PID:G505123

A/Note: parts of this sequence, including the amino ends of three chains isolated from ct

otated below) The mechanism of association with the membrane of the intestinal brush border

C/Complex: Mature enteropeptidase is variously reported to contain two (heavy and light)

ed by a disulfide bond. Possibly, conversion from membrane-bound to soluble forms involve

C/Function:

A>Description: cleaves activation peptide from trypsinogen to produce active trypsin

A/Pathway: intestinal digestive hydrolase cascade

C:Superfamily: enteropeptidase; Clr/Cls repeat homology; LDL receptor ligand-binding rep

C/Keywords: glycoprotein; hydrolase; serine proteinase; transmembrane protein; zymogen

F/22-38/Domain: transmembrane #status: predicted <TM>

F/52-117/Product: enteropeptidase mini chain #status: predicted <MCH>

F/118-799/Product: enteropeptidase heavy chain #status: predicted <HCH>

F/199-999/Domain: LDL receptor ligand-binding repeat homology <LDL1>

F/357-519/Domain: MAM homology <MAM>

F/541-645/Domain: Clr/Cls repeat homology <CLR>

F/658-692/Domain: LDL receptor ligand-binding repeat homology <LDL2>

F/693-798/Domain: scavenger receptor cysteine-rich domain homology #status: atypical <SRCF

F/800-1034/Product: enteropeptidase light chain #status: predicted <LCH>

F/800-1029/Domain: trypsin homology <TRY>

F/116-147, 170, 194, 283, 343, 350, 403, 455, 485, 518, 549, 645, 697, 701, 721, 740, 761, 804, 863, 902, 964

F/787-811, 825-841, 925-992, 956-971, 982-1010/Disulfide bonds: #status: predicted

F/840, 891, 986/Active site: His, Asp, Ser #status: predicted

Query Match 15.5%; Score 724; DB 1; Length 1034;
Best Local Similarity 23.8%; Pred. No. 2e-39;
Matches 258; Conservative 135; Mismatches 330; Indels 360; Gaps 41;

```

QY 55 WVTLAVLVIGLLVLLGIGLVNHLQYRDVRYVK--VKNGYKRTIT---NEKPVAYEN 107
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

Db      21 FTALFALIMVLCAGLIIVSWLTIKGEKDALGKSHARGMTKITSVTYVNIQDRL-- 78
Qy      108 SNSTFVSLASKVDALKLYSGVPLGPRYHKEAAYVTFPSGCVIAY-----W-SFES 160
Db      79 --SVDFKVLADIQIMICEIFOSSMLKNEY-KNSRVLOFENGSVIVIFDLFLAQVNDEN 135
Qy      161 IPOHVEAEKVMKEERVVM----- 180
Db      136 IKEELIQIEANKSSQVAFHIDVNSIDITSELENYSTTSPSTSDKLTSTSPATPGNV 195
Qy      181 ----LP--PRASLSKSFV-----TSVAFPTDSKTORTODNSCSFGIARHAGVELMRF 229
Db      196 SIECLPGSRPCADMLKCIAYDLFCOGELNCPDGSDESKICATACD-----GKRLTES 249
Qy      230 TPGPDSPYF--AHARCONALRGDADSVLSLTFRSDLASCDE-----GSDLV-- 276
Db      250 SGSPDAQYPRKLEASVVCWMIIRVNOGLSIELNFSYFNYSMDVNIYEGGSKILRA 309
Qy      277 -----TVNYTILS----- 283
Db      310 SLMLNPGTIRIPSNQVTVTFILIESDENDYIGFNATYTAFTANSTELANDEKINCPEDEGFC 369
Qy      284 -----PM 285
Db      370 FMIDLANDNEMERIQGTPPTPGPNPDHTFGNASGYIISPTGPGRGQERVGILSLPL 429
Qy      286 EPHALVOLC-----GTYPSYV----- 302
Db      430 EP-TLEPCLSFMYWYGENYKLSINISNDONMEKIIPOKRGNGEMWNNQVTLNETV 488
Qy      303 ---LTFHSSONVILITLTNTERHRRGR-----EATPF-----QLRMSGCG-RLR 345
Db      489 EPKVAFNPKQFLSDIALDISLTGYICNSVLEPFLVPTSPPEL--TDCCGPEFLW 546
Qy      346 KAOGFNSPYRPGHYPNIDCTWNIENNOHVKVRFKFYLLBEPVAGTCPPXUYEI- 404
Db      547 EPNFTTFSMNPNNYPNQAFCWMNLNAQKNGIQHFEFLENIA-----DVVEIR 598
Qy      405 NGEK-----YCGERSQFVYTSNSNKITTFRHSDDSYTDTGLAEVLS---YDSDPC 453
Db      599 DGEEDSDILLAVYTGPGFVEDEVFSTTNMTVLFINDALTYGKGFANFTGYHGLIPPC 658
Qy      454 -PGFTCTGRGCIKREKLCGMDADCTHSDDELNCSGADHQFTCKNKKCKPLFWCDSVN 512
Db      659 KEDNQCNGBCVLLVNLCDGSHCKDSDDEAHCV-----RF-----LN 697
Qy      513 DCGNDSDEGSCSPAQTRCS--NGKCLSKSQCCNGKDC--GDGSDASCP----- 560
Db      698 GTANNSG-----LVQFRIGSIWHTACAEWMTTQTSDDVQLGLGTGNSMPPFSGG 750
Qy      561 ---KVVVVTCTKHTYRCLNGICLSKGNPECDGKED-----CSDGSDKDCGGLRSFT 610
Db      751 GPFVYLTNAP-----NGSLITLTASEQC--FEDSDILILQCNHKS-----CGKKQVA 793
Qy      611 ROA--RVVGGTDADGEMPMQVSLHAGQGHICGASLSLSPMWVLSAAHCYIDDGFRSD 668
Db      794 QEVSPKTYIGANDSRGAMPVVALYTNQ-LICGASLSRMLVLSAAHCYV--RULE 848
Qy      669 PTOWTAFGLHDOSGRASAPVQOERLKRILSHPFENDTTPDYALALELEKPAEYSSWVR 728
Db      849 PSKKKAILGLHMTSLTSPQIVTRLIDELIVNPHNRRKOSDIAMHLEKFAVYTDIQ 908
Qy      729 PICLPDASHVPAGKAIWVGKHTQYGGTALLIQKEIRIVINOTTEBLLPO-QITPR 787
Db      909 PICLPEENQVPPPGICISIAQKGYIYQSSPADILQEDADVPLSNEKQOQOMPENITEN 968
Qy      788 MMCGVFLSGGVDSCGDGGPLSSVEADGRIFGAGVSWMGDCAGRNKRGVYTRPLPRD 847
Db      969 MMCGYEEGGIDSCGDSGGPLMCLL-NNRMLLAGVTSFGIQCALPNRPGVYARVFKTE 1027
Qy      848 WIK 850
Db      1028 WIQ 1030

```

RESULT 3

A43090 enteropeptidase (EC 3.4.21.9) precursor [validated] - bovine

N.Alternate names: enterokinase

C.Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 09-Jul-2004

C.Accession: A43090; A48874; A61436

R.Kikamoto, Y.; Yuan, X.; Wu, Q.; McCourt, D.W.; Sadler, J.E.

A.Title: Enterokinase, the initiator of intestinal digestion, is a mosaic protease compo

A.Reference number: A43090; PMID:94329561; PMID:8052624

A.Accession: A43090

A.Stature: nucleic acid sequence not shown; translated from GB/EMBL/DBJ

A.Molecule type: mRNA

A.Residues: 1-1035 <NIT>

A.Cross-references: UNIPROT:P96072; GB:U09859; NID:G746410; PIDN:AB40026.1; PID:G746411

J.Biol. Chem. 268, 23311-23317, 1993

A.Title: Cloning and functional expression of a cDNA encoding the catalytic subunit of bc

A.Reference number: A48874; PMID:94043122; PMID:8226855

A.Accession: A48874

A.Molecule type: mRNA

A.Residues: 801-1035 <LAV>

A.Cross-references: GB:L19663; NID:G416131; PIDN:AAA16035.1; PID:G416132

A.Note: parts of this sequence, including the amino end of the mature protein, were confi

R.Light, A.; Janska, H.

J. Protein Chem. 10, 475-480, 1991

A.Title: The amino-terminal sequence of the catalytic subunit of bovine enterokinase.

A.Reference number: A61436; PMID:92189715; PMID:11799406

A.Accession: A61436

A.Molecule type: protein

A.Residues: 801-807, 'Y', 809-827 <LIG>

C.Comment: The mechanism of association with the membrane of the intestinal brush border

embrane attachment using a signal-anchor sequence.

C.Comment: Conversion from membrane-bound to soluble forms may involve further processing

C.Complex: mature enteropeptidase is variously reported to contain two (heavy and light)

hide linked

A.Function:

A.Description: cleaves propeptide from trypsinogen to produce active trypsin

A.Pathway: intestinal digestive hydrolase cascade

C.Superfamily: enteropeptidase; Ctr/Cis repeat homology; LDL receptor ligand-binding rep

C.Keywords: glycoprotein; hydrolase; intestine; serine proteinase; transmembrane protein.

F/22-38/Domain: transmembrane #status predicted <TM>

F/52-117/Product: enteropeptidase mini chain #status predicted <MCH>

F/118-800/Product: enteropeptidase heavy chain #status predicted <HCH>

F/199-236/Domain: LDL receptor ligand-binding repeat homology <LDL1>

F/356-520/Domain: MAM homology <MAM>

F/542-647/Domain: Ctr/Cis repeat homology <CTR>

F/659-693/Domain: LDL receptor ligand-binding repeat homology <LDL2>

F/801-1035/Product: scavenger receptor cysteine-rich domain homology #status atypical <SRCT

F/801-1035/Product: enteropeptidase light chain #status predicted <LCH>

F/1801-1030/Domain: trypsin homology <TRY>

F/116,147,170,194,233,263,264,404,456,486,519,550,646,698,722,741,762,864,903,965/Binding

F/786-912,926-842,926-993,957-972,983-1011/Disulfide bonds: #status predicted

F/841,892,987/Active site: His, Asp, Ser #status predicted

Query Match

Best Local Similarity 15.2%; Score 709.5; DB 1; Length 1035;

Matches 251; Conservative 150; Mismatches 320; Indels 365; Gaps 44;

Qy 55 WVTLAVLIGLLVGLIGFLVHMLQYRDV--RVQKVKNGKWRITNENFVDVYNSNST 111

Db 21 FAVLFLVILVLCAGLIIVSWLTIKGEKDALGKSHARGMTKITSVTYVNIQDRLSV 80

Qy 112 EPHVSLASKVDALKLYSGVPLGPRYHKEAAYVTFPSGCVIAY-----W-SFES1POH 164

Db 81 DFKVLADIQIMDDITFQSSNLKNEY-KNSRVLOFENGSIIVIFDLFLDQVSDKNKKEE 139

Qy 165 LVEAEKVMKEERVVM-----LPPRASLSKSFVVTYSV-----V 197

Db 140 LIQIEANKSQTIVTFHIDINSIDITASLENFSTISPAITSEK--LTTSIPLATPGNVS I 197
Qy 198 AFPTDSKTIVORTOONSGSFGHARGVEIM-----RFTTPTGF 233
Db 198 ECPDPSKL-----CADALKCAIADIFCDGELNCPDGSDEKNTCATACDGFLLTGS 249
Qy 234 PDS-----PYPAH---ARCGMALRGADSVLSITFRSFD----- 264
Db 250 SGSEFALHPKPSNNSTAVSCWIRVNOGLSIQNFYFNTYADVAVINIEYEGMGSKILR 309
Qy 265 -----LASCDERGSDDL--TVNYTSLSPMEPHALVOL----- 293
Db 310 ASLWSNNGIIRIIFSNVOTATFLLQSDS--SDYIGFVYTAFAFSKELNNEYKINCNPED 367
Qy 294 --C-----GTYPPSYNLPF-HS----- 307
Db 368 GFCTWIDLDNDNEMERTQGSTFPSTPGPTPDHIFGMSGYISTPFGPGRRRVGLLT 427
Qy 308 -----SONVLLITLITNTERRHGFEATFPQ----- 333
Db 428 LPLDPTPEQACLSPWYMYGENVYKLSINISSDQ--NMEXTIRIQKEGNGQNNYQVLT 484
Qy 334 -----LPMSSCGG 342
Db 485 LNETVERKVSFYGKNOILSDIALDIDISLTYGICNVSVPEPTLVTPPELP--TDCCG 542
Qy 343 --RLRKAQGFNSPYRGHYPPNIDCTWNIIEVNNQHVKRFKFFYLLEPGVPAGTCKPD 400
Db 543 PHDMEWNTTFTSINFPNSYNOAFCLWNLNAQKXIQIHLFQGFLENTA-----D 594
Qy 401 YVEI-NGEKYCGERSQFV-----VTSNSKITVRPHSDQSYTDTGLAEYLS-- 446
Db 595 VVEIRDS-----GDSLSFLAVYTGPGVNDVSTNRMVTVEFTDNMLAKQGFKANFTTGY 651
Qy 447 -YDSSDC-PCQFCRGRGRCIRKELRCDGMADCTDHSDELNC-----SCDAGH--OPTC 496
Db 652 GLGIPCECKEDNFQCKDGECEPIVNLCDGEPHCXGSDAHCVRLFNGTIDSSGLVQFRI 711
Qy 497 KNKFCKPLFWVCDVNDGSDNDEQSCSPAQTPRCNGSKLSKQCGNGKDCGDSDE 556
Db 712 QS-----IMHV-----ACAEN-----WTQIISDVC-----QLLG---LGTNSS 743
Qy 557 ASC-----PRNVVVTCTKATYRCLNGCLSKGNPECDGKEDCSGDSDEKDCD---CGLR 607
Db 744 VPTSTGGGYPVNTAP-----NSSLITLTPSQOC-----LEDSLILLQCNKYSCKK 791
Qy 608 SFTFQA--RVVGGDADGEGWPMQVSLHALGCGHIICGASLSPMWVSAAHCTYDDGFR 665
Db 792 LVTOESVPKIVGSDSREGAWPMVVALYFDQO--QVCGASLVSRLMVSAAHCYVG---R 846
Qy 666 YSDPTQWTAFLGLHDQSRSAFVQVERRLKRIISHPEFNDFTPDYDIALLELEKPAEYSS 725
Db 847 NMESKKAIVGLHMASNLTSPOLETRLLIDIVINPHYNNKRRKNDIAMHLEKNAVYTD 906
Qy 726 MVRPICLPDASHVFPAGKAIWVTGHTQYGGTALILQKEIRVINQTTCENLIPQ-QI 784
Db 907 YIOPICLPEEVOVPPPRICSIAGMGALIIYQGSTADYVLEADVPILNENKCOQOMPEXNI 966
Qy 785 TPRMMCVGLSGVDSQCGSGGPGSLSEVADGRIFGAGVWSWGGCGCKGRKPGYTTPLPL 844
Db 967 TENNVCAGYEKGVDSCQDGSGLMCOE--NNRWLLAGVTSFGYQCALPKRPGVYAAVPR 1025
Qy 845 FRDWIK 850
Db 1026 FTEMIQ 1031

RESULT 4

A56318 enteropeptidase (BC 3.4.21.9) precursor [validated] - human
N:Alternate names: enterokinase
C:Species: Homo sapiens (man)
C:Date: 19-May-1995 #sequence_revision 09-Aug-1996 #text_change 09-Jul-2004

C:Accession: A56318; B43090
R:Klmanoto, Y.; Velle, R.A.; Denis-Keller, H.; Sadler, J.E.
Biochemistry 34, 4562-4568, 1995
A:Title: cDNA sequence and chromosomal localization of human enterokinase, the proteolytic
A:Reference number: A56318; MIM:195234679; PMID:7718557
A:Accession: A56318
A:Molecule type: mRNA
A:Residues: 1-1019 <KIT>
A:Cross-references: UNIPROT:P98073; GB:U09860; NID:G746412; PID:AA50138.1; PID:G746413
R:Klmanoto, Y.; Yuan, X.; Wu, O.; McCourt, D.W.; Sadler, J.E.
Proc. Natl. Acad. Sci. U.S.A. 91, 7588-7592, 1994
A:Title: Enterokinase, the initiator of intestinal digestion, is a mosaic protease compo
A:Reference number: A43090; MIM:194329561; PMID:8052624
A:Accession: B43090
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 749-1019 <KIT>
A:Cross-references: GB:U09860
C:Comment: The mechanism of association with the membrane of the intestinal brush border
ated below) or with amino-terminal myristoylation of the heavy chain.
C:Genetics:
A:Gene: GDB:P857
A:Cross-references: GDB:384083; OMIM:226200
A:Map position: 21q21-21q21
C:Complex: Mature enteropeptidase is variously reported to contain two (heavy and light)
duced by a disulfide bond. Possibly, conversion from membrane-bound to soluble forms involve
ducts.
C:Function:
A:Description: cleaves activation peptide from trypsinogen to produce active trypsin
A:Pathway: intestinal digestive hydrolase cascade
C:Superfamily: enteropeptidase; Clr/Cls repeat homology; LDL receptor ligand-binding rep
C:Keywords: glycoprotein; hydrolase; serine proteinase; transmembrane protein; zymogen
F:1-784/Product: enteropeptidase heavy chain #status predicted <HCH>
F:22-38/Domain: transmembrane #status predicted <TM>
F:184-221/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F:342-504/Domain: MM homology <MM>
F:546-631/Domain: Clr/Cls repeat homology <CLR>
F:643-677/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F:678-783/Domain: scavenger receptor cysteine-rich domain homology #status atypical <SRCF
F:785-1019/Product: enteropeptidase light chain #status predicted <LCH>
F:785-1014/Domain: trypsin homology <TRY>
F:116,147,179,328,335,388,440,470,503,534,630,682,706,725,848,887,909,949/Binding site: c
F:772-996,810-826,910-977,941-956,967-995/Disulfide bonds: #status predicted
F:825,876,971/Active site: His, Asp, Ser #status predicted

Query Match 14.5%; Score 679.5; DB 1; Length 1019;
Best Local Similarity 32.5%; Pred. No. 1.6e-36;
Matches 180; Conservative 80; Mismatches 202; Indels 91; Gaps 20;

Qy 333 QLRPMSSCGG--RLRKAQGFNSPYRGHYPPNIDCTWNIIEVNNQHVKRFKFFYLLEP 390
Db 519 ELF--TDCCGPEFLMEBNTFTSFSTNFPNSYPLAFCWILNAQKXIQIHLFQGFLENT 575
Qy 391 GVPAFCPKDYVEI-NGEK-----YGERGQFVYTSNSKITVRPHSDQSYTDTGFL 441
Db 576 -----INDVIVIRIGEENDSLLAIVYTGPGVNDVSTNRMVTVEFTDNMLAKQGF 628
Qy 442 AEVLS---YDSSDCPG-OPTCGRGRCIRKELRCDGMADCTDHSDELNCSGAGHOF--- 494
Db 629 ANFTTGYHAGTPEPCADHFOCKNGECVPLVNLCDGHLHCEGSDADVC---RFGNG 663
Qy 495 TCRN---KECKPLFWVCDVNDGSDNDEQSCSPAQTPRCNGSKLSKQCGNGKDCD 550
Db 684 TNNNGLVRRFRIQSIW---HTCAEN-----WTQIISDVC-----QLL 719
Qy 551 GDGSDASCPKXNVVVTCTKATYRCLNGCLSKGNPECDG-----KEDSDGS-----D 598
Db 720 GLGSGNSKRIFTST-----DGGPFVXKNTAPDHLITPQOCLDLSRLQCN 768
Qy 599 EKPCDGLSFTQARVNGSTDADEGWPQVSLHALGCGHIICGASLSPMWVSAAHCTY 658
Db 769 HKSGCKLAADITTPKIVGSDSREGAWPMVVALYFDQO--QVCGASLVSRLMVSAAHCY 827

Qy	659	1DDRGFRISDPTQWTAFLGLHDQSRKAPGQERLRKRIISHPPNDPTPYDALLLELE	718
Db	828	YG----RLBESKWTALGLHMKSNLTSPOVPPRLIDRIVINPHYRRKNDIAMMLE	883
Qy	719	KPAEYSSVVRPLCPDASHVPPAGKAIWVGWGHQYGGTGLTIJQKEIRVINQTCEN	778
Db	884	EKAVYTDYIQICLPBENQVPPPRGNCIAGMGVYVYGGTTANIIQEDAVLSNERCQ	943
Qy	779	LLPQ-QITPRMVCVGLSGVDSQGDSDGAPLSSVEADGRIFGAGVSWDGCAGRNKG	837
Db	944	QMPENYITENNICAGYBEGGIDSCQGDSDGFLMCE-INRMFLAGVTSFGYKCALPNRPG	1002
Qy	838	YVTRLPFLFRDWIK	850
Db	1003	YVARVSRETENIQ	1015

RESULT 5

low-density lipoprotein receptor-related protein - mouse
C/Species: Mus musculus (house mouse)
C/Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 09-Jul-2004
C/Accession: JEO315
R/Tomita, Y.; Kim, D.; Magoori, K.; Fujino, T.; Yamamoto, T.T.
J. Biochem. 124, 784-789, 1998
A/Title: A novel low-density lipoprotein receptor-related protein with type II membrane
A/Reference number: JEO315, MUID:96425964, PMID:9756624
A/Accession: JEO315
A/Status: Preliminary
A/Molecule type: mRNA
A/Molecule type: mRNA
A/Residues: 1-1113 <TOM>
A/Cross-references: UNIPROT:Q9Z319; DDBJ:AB013874; NID:g3869144; PIN:BA34371.1; PTD:g31
C/SpeciesFamily: mouse low-density lipoprotein receptor-related protein; LDL receptor ligand
F/337-337/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F/374-408/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F/410-445/Domain: LDL receptor ligand-binding repeat homology <LDL3>
F/447-482/Domain: LDL receptor ligand-binding repeat homology <LDL4>
F/646-682/Domain: LDL receptor ligand-binding repeat homology <LDL5>
F/684-720/Domain: LDL receptor ligand-binding repeat homology <LDL6>
F/723-757/Domain: LDL receptor ligand-binding repeat homology <LDL7>
/869-1097/Domain: tryptsin homology <TRY>

Query Match	14.1%	Score 661.5;	DB 2;	Length 1113;
Best Local Similarity	32.7%	Pred. No. 2.6e-35;		
Matches 158;	Conservative 67;	Mismatches 187;	Indels 71;	Gaps 16;
Qy	428	RHSDQSYTDTPGFLAEYLVSYDSDPCPGQFPCRTGRCIRKELRCDGMDCTDTHSDENLCS	487	
	:	:	:	:
Db	627	QFPESSNQICLLP---NEDVECSPIHFCKRSRCVLSRRCKDQADCDDBDENCG	683	
Qy	488	CDAGHQFTCK-NKFCCKPLFWVCDSVNDGSDNDEQGS-CPAQTFRCSNGKLSKSQOCN	545	
	:	:	:	:
Db	684	CKERALMCEPNNKQCLKHTLLICDGFPPDCPSDMEKNCSFCQDNELCAHECVPRLIMCD	743	
Qy	546	GKDDCGSDSDASCPKVN-----VYICTKTKYCLNG-----LC-----	579	
	:	:	:	:
Db	744	GMVCSQSDSDSDMGCVTLTSKNGNSSSLTVHKSABEHVACADGRETLSQLACKQMGJGEP	803	
Qy	580	-LSKNPECEDEKE-----DCSDGSDDEK-----C--DCGLRSFT	610	
	:	:	:	:
Db	804	SVTLILPEQEGQOMRLVPMNENLNGSTLQGLVLYRRHSPSRSEISLLCSKQDCGRPRPA	863	
Qy	611	R-QARVVGCTADDEGEMPVQVSLNALGQSHCGASLSLSPMWLVSAHCTYIDDRGFRISDP	669	
	:	:	:	:
Db	864	RNNKRIILIGRTSRPRMWCQSLQSEPSGHCQCVLLAKKVVLLVAHCF-BGR---EDA	918	
Qy	670	TQMTAFLLHQQSQSAGVQERLKRIIISHPFDDFPDIDIALLEKPAEYSSMWR	729	
	:	:	:	:
Db	919	DWKVYVEINNLDRHSG-FMQTRPVKTYLLIPRYSRAVVDISVLESLDINETSIVPR	977	
Qy	730	ICLPDASHVPFAGAKIAWVGHTQYGGTGAALLQKGEIRVINOTTCENLLPQO-ITPRM	788	
	:	:	:	:
Db	978	VCLSPSEBYLEPDTCYITGGWH--MGNNKMPFKLOEGSVRIILPEQCSQSYDMKTIITRM	1035	

#	QY	Db	QY	Db
789	MCVFGLSGGVNDCGGDGGPLSSVEADGRIFGAGVSWGDDCAGR-NKPGVYTLPLERD	847		
1036	ICAGYEEGGTVNDCMGDGGPLVCERPGGQWTLFGLTSMGSGVCFSKVLGPGVYSNVSYPVG	1095		
848	WIK 850			
1096	WIE 1098			

RESULT 6

polypeptide - African clawed frog
 C|Species: *Xenopus laevis* (African clawed frog)
 C|Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
 C|Accession: T30337
 R|Yang, J.C.; Lindsey, L.L.; Hedrick, J.L.
 submitted to the EMBL Data Library, March 1998
 A|Description: cDNA cloning of ovochymase, a chymotrypsin-like protease released from *Xer*
 A|Reference number: 220829
 A|Accession: T30337
 A|Status: preliminary; translated from GB/EMBL/DBJ
 A|Molecule type: mRNA
 A|Residues: 1-1524 <YAN>
 A|Cross-references: UNIPROT:Q91674; EMBL:U81290; NID:G2981640; PID:G2981641; PIDN:AC2473
 C|Superfamily: tyrosin related polypeptide; tyrosin homology

Query Match	12.3%;	Score 576.5;	DB 2;	Length 1524;
Best Local Similarity	26.2%;	Pred. No. 1.3e-29;		
Matches 167;	Conservative 89;	Mismatches 205;	Indels 177;	Gaps 23;

```

0Y 228 YPAHARCOMALRGDADSVLSLTFERSFDL--ASCDSRSGDLVAVNTLS-FMEPHALVOL 293
Db 333 YSINSVCMMMLAVOKAKTIEIRFLODIEDHATC-----TFDLSTFTVEKMIKRV 383
0Y 294 CGTVPSPYNLTFHSSONVLLITLINTERRRHGFEATFPOLP--RMSSCG-GRLRKAQGT 350
Db 384 CQSTIPSPILV---RSNKVYTVTFPSDGTFTGGRGFIEQIOLAIPTKASACCSAKILKKKM 440
0Y 351 FNSPFPYBGHYPPNIDCTWNIEVPNNQHVVRKRFYLLBGRVAG--TCPPDYEI-NG-- 406
Db 441 IYSPVPPPYPLPKTCSWITAEPEHIVLKEEDF-----NVEYGGCIYDAEYVDGAE 495
0Y 407 -----EKYGEKSGQFVTVSNSNKITVRFFHSDSYTDGTGLAEYLSYSDSPCEQPTCRT 461
Db 496 EKQILARICQYTLPLPISPSPENTMLIRFKTDMENSYPGFKVKK----- 538
0Y 462 GRCIRKEIRCDGEMADCTDHSDELNCSADAGHQFTCKNK-FCKRLFWVCDSVNDGDSNDE 520
Db 539 -----GFVPEKQFSLPV----- 551
0Y 521 QGSCGPAQTFRCSNCKLSKSGCQCKDGDGSDSDASCPKVVVYTCIKATYRCLNGLCL 580
Db 552 -----DD-----TFPISML----- 560
0Y 581 SKNPECCDKEDCSDSGDEKDCDCCGLRSFTRO--ARVYGGTDADGEMWQVSLHALGQ 637
Db 561 --HERATALD-----VCGWAPMTPKWMLPRIIVGGEBASPNSWMOVOIFELRT 606
0Y 638 GHICGASLIPMWLWLSAAHCYIDRGFRSYDPTQMTAFGLHPOQSRAFGVERRLKRI 697
Db 607 FH-CEGAIISPMIITPAHC-----IRAEBSYTVIADNHRMLNEST-EQIRNIXTI 658
0Y 698 ISHPFNDFTEDYDIALLELEKPAEYSSWVRPICLPDASHVFPAGKAIWYTGHTQYGG 757
Db 659 RIHDVYNSEYNDIALLYLEBPLDLNDVVRVPCJPEBESVLTPASVCVVTGNGMTVEDG 718
0Y 758 TGAIILOGEIRVINQTCCE-MLLPQOITPRMWCVGFLSG-GVDSOCGDSGGGFLSSVEAD 815
Db 719 QPALGLQOLQPLILDSIICNTSYSGELTDHMLCAGFPBSKEDACGDSGGGFLVQCNEX 778
0Y 816 GRIFGAVVSWDGCAGRRKPGVYTRLLPLFRDMIKENT 853

```

Db 779 EGFSTYGLVSWGEGGVSKPGVYTKVRLFTWI-QNT 815

RESULT 7

154763

Ra-reactive factor (EC 3.4.21.-) 1 precursor - human

N/Alternate names: mannose binding protein-associated serine proteinase 1 (MASP-1)

C/Species: Homo sapiens (man)

C/Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 09-Jul-2004

C/Accession: 154763; JN0883

R/Sato, T.; Endo, Y.; Matsushita, M.; Fujita, T.

Int. Immunol. 6, 665-669, 1994

A/Title: Molecular characterization of a novel serine protease involved in activation of

A/Reference number: 154763; MUID:94289349; PMID:8018603

A/Accession: 154763

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: mRNA

A/Residues: 1-699 <SAT>

A/Cross-references: UNIPROT:P48740; GB:D28593; NID:g790963; PID:BAA05928.1; PID:g471128

R/Takada, F.; Takayama, Y.; Hattuse, H.; Kawakami, M.

Biochem. Biophys. Res. Commun. 196, 1003-1009, 1993

A/Title: A new member of the C1s family of complement proteins found in a bactericidal

A/Accession number: JN0883; MUID:94059062; PMID:8240317

A/Accession: JN0883

A/Molecule type: mRNA

A/Residues: 1-234,'E',236-284,'G',286-498,'K',500-542,'K',544-642,'S',644-699 <TRAK>

A/Cross-references: DDBJ:D17525; NID:9439712; PID:BAA04477.1; PID:g439713

A/Experimental source: liver

C/Comment: This is a serum bactericidal factor that activates complement C4 and C2 compo

C/Genetics:

A/Gene: GDB:MASP1, GDB:CRARF, CRARF1, PRSS5, MASP

A/Cross-references: GDB:361104; GDB:330954; OMIM:600521

A/Map position: 3q27-3q28

C/Superfamily: complement-activating serine proteases C1r/C1s/MASP; C1r/C1s repeat homol

C/Keywords: beta-hydroxyasparagine, complement pathway; duplication, glycoprotein; hydroc

F:1-17/Domain: signal sequence #status predicted <SIG>

F:18-448,449-699/Product: Ra-reactive factor #status predicted <MAT>

F:19-135/Domain: C1r/C1s repeat homology <C1R1>

F:143-181/Domain: BGF homology <BGF>

F:185-294/Domain: C1r/C1s repeat homology <C1R2>

F:301-362/Domain: complement factor H repeat homology <FHL>

F:367-432/Domain: complement factor H repeat homology <FHL>

F:449-691/Domain: trypsin homology <TRY>

F:49,118,407/Binding site: carboxylate (Asn) (covalent) #status predicted

F:73-91,143-157,153-161,168-181,185-212,242-260,301-349,329-362,367-414,397-432,436-572,

F:159/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted

F:448-449/Cleavage site: Arg-ile (autolytic) #status predicted

F:490,552,646/Active site: His, Asp, Ser #status predicted

Query Match 12.2%; Score 570; DB 1; Length 699;

Best Local Similarity 25.3%; Pred. No. 1.5e-29;

Matches 193; Conservative 111; Mismatches 268; Indels 192; Gaps 32;

Qy 203 SKTYQRPQDSCSGFARQVLEMRFTTPGPDPSPYAHARCMALRGDADSVLSTFRS 262

Db 15 SKSAHATVEILANNFG-----QIOSPGYPS-YPDSDEVNTITVPDGRFKLYFMH 64

Qy 263 FDLASCDERSDLVTYVNTLSPMEPHALVOLCG-----TTPPSY-NLTFH 306

Db 65 FNLSSVLCGYDYKV-----ETEDQVLAIFCCGFTTDTQTPQGEVVLSPGSMSTFR 119

Qy 307 SSONVLLITLITTERHPGFATFPOLP----- 335

Db 120 SD-----FSNEERFTGFDAYMAVDVDECKEREDEELSCDHYCHNYIGYVCSGRF 170

Qy 336 -----RMSSCGRLRKAQGTENSPYRGHYEPNDCTMNIEVPNNQHYKVFKEF 364

Db 171 GYLILHNDNRCKRCEGSDNLFTQRTGVTSPFPNPKYKSECLYTLIEEGFMVNLQPED 230

Qy 385 FYLLLE--PGVPAGTCPPDYVEIN-GEK-----YCGERSQFVNTSNKILTVRFHSDQSYTD 437

Db 231 IFDIQDHPEVP---CPDYDIKIKGPKVLPFGCEKAPPEPISQSHSVLLIFHSDNAEN 267

Qy 438 TGFALAYLYSDSSDCPP-----GQFCTGTGCRIRKEKRCGDMADCTHSDLNSCDA 490

Db 288 RGMRLSYRA--AGNECEPELOPVPHGKIBPSQAKYFPK-----DQVLVSCDT 331

Qy 491 GHQP-----TCKRKFKCPPLFW-----VCSYVNDCCGNSDQGGSCPAQ-----TFR 531

Db 332 GYVVLKDNVEMDFQIECLNDGTWNSKIPCTKIVD-----CRAPBLEHGLITFS 381

Qy 532 CSNGKCLSKSQCGNGKDDCGDSDSEASCPEKVVNTCTKHTYRC-----LNGCLSKGNP 585

Db 382 TRNNLTYYSEI---KYSCOE-----PYKMLNNNTGYLTSAGQWMNKY-LGRSLP 430

Qy 586 ECDGKEDCSDSDKDCDCCGLRSFTQ--ARVVGTDADGEMPMOWSLHALGQGH 643

Db 431 TCLPV-----CGLPFKSRKLMARIFNGRPAKGTTPWJAMLSHLNGQFPCCG 477

Qy 644 SLTSPMIVLSAACHY---SD--DRGRVAD---PTQWTAFLGHDQSQSAPEVQRRRK 695

Db 478 SLTSSMIVYLAACHQSLDPEGPTLRDSDLSPSDFKIIIGKH-WRLNSDENEQHLGVK 536

Qy 696 RIISHPEFNDFTPDYDIALLEKPAEYSGMVRPCLPDASHVPAGKAIWYTGWHTQY 755

Db 537 HTLHHQYDPTNFENDVALVELLESPLNAFWPILCPBGPQ--QSGANVYISGWC-KQF 593

Qy 756 GGTGALLDKGEIRVINTOTTENL---LPQQTIPRMVCYGLSGVDSGQDSGGFLSV 812

Db 594 LQRFPELTMEIEIPIDHSTQCKAVAPLKKKTVRDMICAGEKGGKADACGSDGPMVLT 653

Qy 813 EAD-GRIFGAGVYWGSDGAGRRKPGVYTRPLFRFMWIKENTGV 855

Db 654 NREGQVLYLVGYSWGDDGCKKDRYGVYISYIHNKMKIQRVTVG 697

RESULT 8

KOMSP1

Plasma kallikrein (EC 3.4.21.34) precursor - mouse

C/Species: Mus musculus (house mouse)

C/Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 09-Jul-2004

C/Accession: A36557

R/Seidah, N.G.; Sawyer, N.; Hamelin, J.; Mion, P.; Beaubien, G.; Brachpapa, L.; Rochement

DNA Cell Biol. 9, 737-748, 1990

A/Title: Mouse plasma kallikrein: cDNA structure, enzyme characterization, and compari

A/Reference number: A36557; MUID:91090844; PMID:2264328

A/Accession: A36557

A/Molecule type: mRNA

A/Residues: 1-638 <SEI>

A/Cross-references: UNIPROT:P26262; GB:M58588; NID:9200358; PID:AAA63393.1; PID:9200359

A/Note: part of this sequence, including the amino ends of both the heavy and light chain

C/Comment: This protein, synthesized in the liver, circulates as a noncovalent complex w

C/Comment: are linked by one or more disulfide bonds.

C/Superfamily: coagulation factor XI; trypsin homology

C/Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; infla

F:1-19/Domain: signal sequence #status predicted <SIG>

F:20-330/Product: plasma kallikrein heavy chain #status experimental <HCH>

F:20-109/Domain: apple repeat <AP1>

F:110-199/Domain: apple repeat <AP2>

F:200-289/Domain: apple repeat <AP3>

F:291-380/Domain: apple repeat <AP4>

F:391-638/Product: plasma kallikrein light chain #status experimental <LCH>

F:391-621/Domain: trypsin homology <TRY>

F:21-104,47-77,51-57,111-194,137-166,141-147,201-284,227-256,231-237,292-375,318-347,322-

F:127,215,308,396,494/Binding site: carboxylate (Asn) (covalent) #status predicted

F:434,483,578/Active site: His, Asp, Ser #status predicted

Query Match 11.3%; Score 528.5; DB 1; Length 638;

Best Local Similarity 27.2%; Pred. No. 6.8e-27;

Matches 164; Conservative 85; Mismatches 232; Indels 175; Gaps 32;

Qy 271 RGSDDLVTYVNTLSPMEPHALVOLCGTYPSPYVLTFTSSONVLLITLITTERHPGF--E 328

Db 32 RGSDDLAAITTP-----DAQYCSQKSKCTHPRCLFSF-----LAVTPRKETNKRFGCFMKE 81

```

Qy 329 ATFFOLPRMSSCG---GLRKAQGTENSPYRGHYP-----NIDCTWNIE-----V 372
      |||
      82 SITGTLPRHRTGALISGSHLKQCGHQSACHRDYIKGLDMRGSNFNISKTNIEBCQKLC 141
Qy 373 PNNQHVKKRFR---FYLLF-----PGVPACT----- 396
      |||
      142 TNNHFCQF--FTYATSAFPRPEYRKCKLKHASGTPITSKADNLVSGFSLKSCALSEIG 200
Qy 397 CPKQ-----YVEINGEKYCGERS---QFVNTSNKNTTVAFHDQSTYDQ---FLAE 443
      |||
      201 CPMDIFQHSADNLVNSQVITPDPAVCCTICTFPHNCLFTFTYNNWETESQRNVCFLKT 260
Qy 444 YLSYDSPDPCPGQ-----FTCSTGR---CIRKELRCQDADCTHDSBLNCSADAGH 492
      |||
      261 SKSGRPSPIRPOENALISGYSLLTCKRKTRPCHSKL-----YSGVDPEGELNATFVQGA 315
Qy 493 ---OFTCKNRFCKPLFWVCDVSV-NDCGNSDEQCGSCPAQTFRCNGKCLSKSQCCNGKD 548
      |||
      316 DVCQETC-TKTRICQCFIYSLLPQDC---KEEGCKC---SLRLST----- 353
Qy 549 DCGGSDASCPKXNVVTCGHTYRCLNG-----LCLSKNPREDCGKEDSGSDGDKDC 602
      |||
      354 ---DQSP-----TRITYGQSSGYSRLCLVDSPOCTTKIN----- 388
Qy 603 DCGLRSPTRQARVVGTDADGEMPMQVSLHA--LGQGHICGASLISPMVLVSAHCYID 660
      |||
      389 -----ARIVGNTNALSIGEMPMQVSLQVLSQHLGCGSLTIGQWLTAAHCF-- 436
Qy 661 DRGFRYSPTQWTAFLGLHDQSQ--RSAPGVQERLKRITISHPPNDFTFYDIALLELE 718
      |||
      437 -DGIPY--PDVWRITYGILSLSEITKEP---SSRIKELIHQEVKVEGNDIALIYKQ 490
Qy 719 KPARYSSWVRICLPDASHVFPAGKALVWGMGHQVYCGTALLQKEILVINQTCEN 778
      |||
      491 TPLNLTBERQKICLPKSKADNTTITNCTVWGMGYTKEGSETQNLQKATITLVPNEBCQK 550
Qy 779 LLPQO--ITPRMAMCVGFLSGVDSCQDGGCLPSVEADGRIFGAGVSWMGDCAGRNKPG 837
      |||
      551 KYRIVYINKQMICAGYKGGTADCKGDSGGLP-VCKHSGRMQVIGITWSGEGCCRKQDGP 609
Qy 838 VYTRLPLEFRMNIKENT 853
      |||
      610 VYTKVSEYMDWILEKT 625

```

RESULT 9

KORTPL

Plasma kallikrein (EC 3.4.21.34) precursor - rat

N:Alternate names: Fletcher factor; kininogenin; serum kallikrein

C:Species: Rattus norvegicus (Norway rat)

C>Date: 30-Sep-1992 #sequence revision 30-Sep-1992 #text_change 09-Jul-2004

C:Accession: A39180; A33320; S06851; I53041; S06852

R:Reubien, G.; Rosinski-Chapin, I.; Mattei, M.G.; Mbikay, M.; Chretien, M.; Seidah, N.G. Biochemistry 30, 1628-1635, 1991

A:Title: Gene structure and chromosomal localization of plasma kallikrein.

A:Reference number: A39180; MUID:91129236; PMID:1993180

A:Accession: A39180

A:Molecule type: DNA

A:Residues: 1-638 <BEA>

A:Cross-references: UNIPROT:P14272; GB:J05315

A>Note: the authors translated the codon GAG for residue 81 as Gln

R:Seidah, N.G.; Lidenheim, R.; Mbikay, M.; Hamelin, J.; Lutfalla, G.; Rougeon, F.; Lazure DNA 8, 563-574, 1989

A:Title: The cDNA structure of rat plasma kallikrein.

A:Reference number: A33320; MUID:90091743; PMID:2598771

A:Accession: A33320

A:Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-638 <SEI>

A:Cross-references: GB:M0282; NID:9205010; PIDN:AAA4163.1; PID:9205011

A>Note: part of this sequence, including the amino ends of both the heavy and light chain

R:Paquin, J.; Benjannet, S.; Sawyer, N.; Lazure, C.; Chretien, M.; Seidah, N.G. Biochim. Biophys. Acta 999, 103-110, 1989

```

A:Title: Rat plasma kallikrein: purification, NH(2)-terminal sequencing and development
A:Reference number: S06851; MUID:90089457; PMID:2597701
A:Accession: S06851.
A:Molecule type: protein
A:Residues: 20-45;391-413 <PAQ>
R:Seidah, N.G.; Lidenheim, R.; Mbikay, M.; Hamelin, J.; Lutfalla, G.; Rougeon, R.; Lazure DNA Cell Biol. 8, 563-574, 1989
A:Title: The cDNA structure of rat plasma kallikrein.
A:Reference number: I53041
A:Accession: I53041
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-638 <RRS>
A:Cross-references: GB:M58590; NID:9206721; PIDN:AAA42069.1; PID:9206722
C:Comment: This protein, synthesized in the liver, circulates as a noncovalent complex w/
C:Comment: The zymogen is activated by factor XIIa, which cleaves the molecule into a 115
are linked by one or more disulfide bonds.
C:Genetics:
A:Gene: PK
C:Superfamily: coagulation factor XI; trypsin homology
F:1-19/Domain: signal sequence; duplication; fibrinolysis; glycoprotein; hydrolase; infla
F:20-350/Product: plasma kallikrein heavy chain #status experimental <MAT1>
F:20-109/Domain: apple repeat <AP1>
F:110-199/Domain: apple repeat <AP2>
F:200-289/Domain: apple repeat <AP3>
F:291-380/Domain: apple repeat <AP4>
F:391-638/Product: plasma kallikrein light chain #status experimental <MAT2>
F:391-621/Domain: trypsin homology <TRY>
F:21-104,47-51,57,111-194,137-166,141-147,201-284,227-256,231-237,292-375,318-347,322-
F:127,215,308,453,459,494/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:396/Binding site: carbohydrate (Asn) (covalent) #status experimental
F:433,483,578/Active site: His, Asp, Ser #status predicted

```

Query Match 10.8%; Score 507.5; DB 1; Length 638;

Best local similarity 26.5%; Pred. No. 1.6e-25;

Matches 183; Conservative 79; Mismatches 220; Indels 209; Gaps 35;

```

Qy 271 RGSGLVYVNTLSMEBHALVQCGTPSPSYNLTFFHSQVNLITLTNTERHRRPGF--E 328
      |||
      32 RGSGLVAIY---PDAGHC-QKQCTFPRCLFSF-----LAVSPKXETDKRGCKMKE 81
Db
Qy 329 ATFFOLPRMSSCG---GLRKAQGTENSPYRGHYP-----NIDCTWNIE-----V 372
      |||
      82 SITGTLPRHRTGALISGSHLKQCGHQSACHQDIYBSLDMRGSNFNISKTDSEBCQKLC 141
Db
Qy 373 PNNQHVK--VREKFYLLF-----PGVPACTCPDYVEING--EKYCG----- 411
      |||
      142 TNNHFCQFPTYATGAFPRPEYRKCKLKHSSGTPITSKADNLVSGFSLKSCALSEIG 200
Qy 412 -----ERSQFVNTSNKNTTVAFHDQSTYDQFLAEYLSYSSDPCPGQFCRTGRCI 465
      |||
      201 CPMDIFQHSADNLVNSQVITPDPAVCCTICTFPHNCLFTFTYNNWETESQRNVCFLKT 229
Qy 466 KEELRCQMGADCTHDSBLNCSADAGHQTCKNFKCKPLF-----WVCDVSV-NDCGDN 517
      |||
      230 -----VCTFHP--NC-----LFTTYNNWETESQRNVCFLK 259
Db
Qy 518 SDEQCGSCPA-----QTFRCNGKCLSKSQCCNGGDDCG-----DGS 554
      |||
      260 TSKSGRPSPIRPOENALISGYSLLTCKRK-----ARPECHFIYGVAFEGELNATFVQGA 315
Qy 555 D--EASGPXNVVTCGHTYRCLNGCLSKGNPECGKEDCS-----DGS----- 597
      |||
      316 DACQETCTK--TIRQCFYSLLPQDCKAG--C-----KCSJRLSTDGSPTRYTYEAQG 366
Db
Qy 598 -----DEKDCGLRSFTROARVVGTDADGEMPMQVSLHA--LGQGHICGA 643
      |||
      367 SSGYSALCKVYBSSDCTTKI-----NARIVGNTNALSIGEMPMQVSLQVLSQHLGCG 421
Qy 644 SLSPNVLVSAHACYIDRGRYSPTQWTAFLGLHDQSQ--RSAPGVQERLKRITISHP 701
      |||
      422 SIIRQWILTAAHCF--DGIPY--PDVWRITYGILSLSEITKTP---FSSIKELIHQ 473

```

Qy 702 FENDFTDYDIALLELEKPAEYSSWVRPICLPDASHVFPAGKAIWVGHTOYGTGAL 761
 Db 474 KYKSSGSDYIALKLTQPLNTYEFQKPCPSKADNTNTITNCTWVGWGTTKRGSTQN 533
 Qy 762 ILQGEIRVINTQTCENLFPQO-ITPRMVCVGLSGVDSCQDGGPPLSVEADGRIFG 820
 Db 534 ILQKATPLPVNEBCQKRYRDYVITKQWICAGYKGGIDACKDGGSGPL-VCKHSGRWOL 592
 Qy 821 AGVSWGDGCGARKKPGVYTRLPFRDMKE 851
 Db 593 VGITSGEGCARKQPGVYTKVAEYIDWILE 623
 RESULT 10
 PLPG
 plasmin (EC 3.4.21.7) precursor - pig (fragment)
 N/Alternate names: plasminogen
 N/Contains: miniplasminogen
 C/Species: Sus scrofa domestica (domestic pig)
 C/Date: 07-Sep-1990 #sequence revision 01-Nov-1996 #text_change 09-Jul-2004
 C/Accession: S03733; S03737; A25834
 R/Schaller, J.; Marti, T.; Roesele, S.J.; Kaempfer, U.; Rickli, E.E.
 Fibrinolysis 1, 91-102, 1987
 A/Title: Amino acid sequence of the heavy chain of porcine plasmin. Comparison of the ca
 A/Reference number: S03733
 A/Accession: S03733
 A/Molecule type: protein
 A/Residues: 1-560 <SCH>
 A/Cross-references: UNIPROT:P06867
 R/Brunsholz, R.A.; Lerch, P.G.; Schaller, J.; Rickli, E.E.; Lergier, W.; Manneberg, M.;
 Eur. J. Biochem. 114, 465-470, 1981
 A/Title: Comparison of the primary structure of the N-terminal CNBr fragments of human,
 A/Reference number: S03735; MUID:81212097; PMID:7238497
 A/Accession: S03737
 A/Molecule type: protein
 A/Residues: 1-57 <BRU>
 R/Marti, T.; Schaller, J.; Rickli, E.E.
 Eur. J. Biochem. 149, 279-285, 1985
 A/Title: Determination of the complete amino-acid sequence of porcine miniplasminogen.
 A/Reference number: A25834; MUID:85203907; PMID:3846533
 A/Accession: A25834
 A/Molecule type: protein
 A/Residues: 450-790 <MAR>
 C/Function:
 A/Description: dissolves the fibrin of blood clots; acts as a proteolytic factor in a va
 A/Pathway: fibrinolysis
 C/Superfamily: plasmin; kringle homology; plasminogen-related protein precursor homology
 C/Keywords: fibrinolysis; glycoprotein; hydrolase; kidney; kringle; plasma; serine prote
 F/1-790/Product: plasminogen #status predicted <PRO>
 F/1-77/Domin: plasminogen-related protein precursor homology (fragment) <PLPH>
 F/1-77/Domin: activation peptide #status predicted <APT>
 F/78-560/Product: plasmin chain A #status predicted <ACH>
 F/84-162/Domin: kringle homology <KR1>
 F/166-243/Domin: kringle homology <KR2>
 F/256-333/Domin: kringle homology <KR3>
 F/358-435/Domin: kringle homology <KR4>
 F/450-790/Product: miniplasminogen #status experimental <MIN>
 F/461-540/Domin: kringle homology <KR5>
 F/561-790/Product: plasmin chain B #status experimental <BGH>
 F/561-783/Domin: trypsin homology <TRY>
 F/30-54,34-42,84-162,105-145,133-157,166-243,169-297,187-226,215-238,256-333,277-316,305
 bonds: #status predicted
 F/602,645/Active site: His, Asp, Ser #status predicted
 Query Match 10.8%; Score 506.5; DB 1; Length 790;
 Best Local Similarity 23.2%; Pred. No. 2.4e-25;
 Matches 202; Conservative 100; Mismatches 301; Indels 267; Gaps 36;
 Qy 106 ENSNSTFVSLASVYKDAL--KLTVSGVPLGPGYHKSNAVTSSEGSVIAYVWSEFS-- 160
 Db 59 ENSKT---SFIAMRWVLFKRIYLSCKTGNKGYRGTTSKTKSGVICQKMSVSSPH 114

Qy 161 IPOHVEAEERVAERVWMLPPRARSLSFVTSVVAPEPDSKTYQRTODNSCSFGLHA 220
 Db 115 IPRYSKEKPLAGLLEENYCNPNDEK-----GPMCTTDBET----- 152
 Qy 221 RGVLEIRFTTGGPPDPYPAHACQWALRGDADSVLSLFRSDLASCDERGDLYTVYN 280
 Db 153 -----RFYCDIPE-----CEDCMHSGEHYKGLSKTWSGLE-----COSWSG----- 192
 Qy 281 TLSPEPHALVOLCGTYPPSYNLTFHSSQVLLITLITYTER-----HPGEATFOL 334
 Db 193 -----QSPHAGYLPKSF-PKALIKMYCN-----PDGEPPKCFPTDPNRMERFCDI 240
 Qy 335 PRMS-----CGGLRRAQGTFN-----SPYYPGHYPNIDCTWN 369
 Db 241 PRCTTPPTSGPTTGYQCLKRGGENYRGTVASGHTCQWSAQSPKHNRTPENPCK-N 299
 Qy 370 IEV-----PNNQVY-----KYRKFPYL-----LEGVAGTGP--KDYV 402
 Db 300 LLENNYCRNPDGETAPWCYTTSSEVRWDYCKIPCGSITSTEHLDAVPPEQTPVAQDCY 359
 Qy 403 EINGEKYCGERSQFVVTNSNK-ITVRFSDD-----SYDTGFLAYLSYSSDPGPGF 457
 Db 360 KMGESTRGTSITITGRKQGSVSMTPRHHEKTPGNFPAGLITMYCNRPADKSPWCY 419
 Qy 458 T-----CRTRGCRKE----- 468
 Db 420 TTDPRVRWEYCNLKKSETEQVTNFPALAQVSVEDLSBDCWFGNGKRYRGKATTVAG 479
 Qy 469 LRCDGADCTDHSDELANCSDAGHOFTCKNPKFLFWCDSDVNDGDNDSDEGGSC-- 525
 Db 480 VPQEWAAQEPHNSI-FTPETNPRAGLEKNYCR-----NPGDDGMPWCYTT 526
 Qy 526 -PAQTF-RCNGKCLSKSQDCNGDGGSDASCPKVVVYCTKHTYRCINGLCLSKG 583
 Db 527 NPQKLPDYCDVPCVTS-----FDCK-----PKV----- 552
 Qy 584 NPECDGKEDSDSDKDCDCCGLRSFTROARVVGTDADGEMPVQVSLHALGQHI CGA 643
 Db 553 -----EPKCK-----PARVVGCGVSIIPHSWPMQISLRYRGRHFCG 589
 Qy 644 SLTSPWLVSAAMCYIDDGFRYSPTQWTAFLGLDQGRSA.PGVQERRLRKIIISHPF 703
 Db 590 TLISPMVVLTKACL-----EKSSPSSYKVLIGAEH-YHLGEGVQEDIVSLFYEP-- 641
 Qy 704 NDFTPYDIALLELEKPAEYSSWVRPICLPDASHVFPAGKAIWVGHTOYGTGALIT 763
 Db 642 ----SEADIALKLSSPAVITDKVTPACLPTRYVYVADRTACTITGMGRTK-CTYGAGL 696
 Qy 764 QKGEIRVINTQTCENLFPQOITPRMVCVGLSGVDSCQDGGPPLSVEADGRIFGA 821
 Db 697 KEARLPVIEKVCNRYEYLGKVSFNELCAGHLAGIDSCQDGGPPLVCFEKDYIL-Q 755
 Qy 822 GVSWDGCGARKKPGVYTRLPFRDMKE 851
 Db 756 VGITSGEGCARKKPGVYTRVRSFTWIB 785
 RESULT 11
 S15468
 Complement C3b/C4b inactivator (EC 3.4.21.-) precursor - African clawed frog
 N/Alternate names: C3b/C4b inactivator factor I
 C/Species: Xenopus laevis (African clawed frog)
 C/Date: 20-Feb-1995 #sequence revision 20-Feb-1995 #text_change 09-Jul-2004
 C/Accession: I51601; S15468
 R/Kunnath-Muglia, L.M.; Chang, G.H.; Sim, R.B.; Day, A.J.; Ezekowitz, R.A.
 Mol. Immunol. 30, 1249-1256, 1993
 A/Title: Characterization of Xenopus laevis complement factor I structure--conservation
 A/Reference number: I51601; MUID:94019415; PMID:7692240
 A/Accession: I51601
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: mRNA
 A/Residues: 1-613 <KUN>

C:Superfamily: serine proteinase stubble-stubloid: trypsin homology
C:Keywords: hydrolase, serine proteinase, transmembrane protein
F:61-77/Domain: #status Predicted <TM>

F:391-621/Domain: tryptophan homology <TRY>
F:211-104-47-77,51-87,111-194,137-166,141-147,201-284,227-256,231-237,292-375,332-338,363
F:127,306,336,453,494/Binding site: carboxyterminal (Asn) (covalent) #status experimental

F,318-347,340-345/Disulfide bonds: #status predicted
F,390-391/Cleavage site: Arg-116 (coagulation factor XIIa) #status predicted
F,434,483,578/Active site: His, Asp, Ser #status predicted

Query Match 10.6%; Score 495; DB 1; Length 638;
Best Local Similarity 26.8%; Pred. No. 1e-24;
Matches 183; Conservative 70; Mismatches 209; Indels 222; Gaps 31;

```
QY 328 EATPF-OLPMSSCGGRLRAKQGF-----NSPYGHYPNIDCTWIE-----VPN 374
DB 6 QATFISLEFATVSGGCLTQLYENAFPGGVASNTYNAQYCMKCTFHRCILFSLPA 65
QY 375 NQHVKAFKFFYLPEVPAGTCPKDY---VEINGEKYCGERSQV-----V 418
DB 66 SSINDMEKRFEGCFKDSV-TGTLPEKVRTGAVSGHSLKQGHQISACHRDYKGVDRGV 124
QY 419 TSNMNTK-----TYRHSDQ-----SY 435
DB 125 NFNVSKVSVSEBCKRCTNNIRCFSTATQTFHKAERYNNCLIKYSPGCTPAIKVLSN 184
QY 436 TDNGF-----LAELYSDSSDPC-----PGQFCTGRGICRKLKCDGMAD 476
DB 185 VESGFLSLKPCALSIGGHMNI FQHLAFSDVDVARVLTPEAFVCKT-----I 230
QY 477 CTDSDELNCSGADHGFCKKFKCPF-----VDCSV-NDGSDNDEOG---CSC 525
DB 231 CTYHP--NC-----LFFTYTNWKIESQRVCLIKTISESGTPASSST 270
QY 526 PAQRF-----RCSNGKCLSKSQOCCGKDCGSDGSEASCPKNNV--TCTK--- 569
DB 271 PQENTISGYSLLTCKRKLPEPCHSKIYP--GVDFGSELNVTYKGVNVOETCTKMRIC 328
QY 570 --HYRCLNGLCLSKNGPECDGKEDCSGSDKDCDGLR-----SFTROA----- 613
DB 329 QEFFYSL-----PEDCK---EKKCKFKLMSMGSTRIAYGTQSSSGV 370
QY 614 -----RVGGTDADEGEMVQVSLAH--LGQGHICGASLSPNVLV 652
DB 371 SLRLCNTGDSVCTTKTSTRIIVGCTGSSMGMPQVSLQVLTQORHLCGSLILGHQWL 430
QY 653 SAACHYIDDRGFRSDPTQMTAFILGLHDQO--RSAAGVERRLKRIISHPFENDPFEDY 710
DB 431 TAAHCF--DGLPIQD--VWRIYSGLINLSDITQTPSQ--IKELIITHONYKVSNGH 482
QY 711 DIALLELEKPAEYSSMVRPICLPDASHVFPAGKAIWTVGWHQYGGTGALLIQGEIRV 770
DB 483 DIALIKLQALNLTPEPKPICLPBPKDPTSTIYNCWTVGWFSGEKEIQLNIQKXNIP 542
QY 771 INOTTGENTLPQ-OITPRMNCVGLSGVDSCQDSCGGLSVSEADGRIFGAGVSWGQ 829
DB 543 VTNEBCKRKYQDYKITQRMVCAQYKEGKDKACKGDSGGL--VCKHNGMRLVGITSWGEG 601
QY 830 CAGRNKPGVYTRLLPFRDWIKENT 853
DB 602 CARBQPGVYTKVAEYMDWILEKT 625
```

RESULT 14

B61545
Plasma (EC 3.4.21.7) precursor - sheep (fragments)
N/Alternate names: Plasma
N/Contains: miniplasminogen
C/Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
C/Date: 28-Oct-1994 #sequence_revision 01-Nov-1996 #text_change 09-Jul-2004
C/Accession: B61545; S28200
R/Schaller, J., Rickli, E.E.
Enzyme 40, 63-69, 1988
A/Title: Structural aspects of the plasminogen of various species.
A/Reference number: A61545; PMID:89005015; PMID:3168975
A/Accession: B61545
A/Molecule type: protein
A/Residues: 1-37,38-117 <SC>
A/Cross-references: UNIPROT:P81286

R/Schaller, J., Straub, C., Kaempfer, U., Rickli, E.E.
Protein Seq. Data Anal. 5, 21-25, 1992
A/Title: Complete amino acid sequence of ovine miniplasminogen.
A/Reference number: S28200; PMID:93149955; PMID:1492092

A/Accession: S28200
A/Molecule type: protein
A/Residues: 118-460 <SC>
C/Superfamily: plasmin; kringle homology; plasminogen-related protein precursor homology;
C/Keywords: fibrinolysis; glycoprotein; hydrolase; kringle; plasma; serine proteinase; z
F,1-37,38-117,118-460/Product: plasminogen (fragments) #status experimental <PRO>
F,1-37/Domain: activation peptide (fragment) #status experimental <AP>
F,38-117,118-230,231-460/Product: plasmin (fragments) #status experimental <MAT>
F,41-118/Domain: kringle homology <KR4>
F,118-460/Product: miniplasminogen #status experimental <MIN>
F,132-211/Domain: kringle homology <KR5>
F,226-460/Domain: plasmin chain B #status experimental <BCH>
F,221-453/Domain: trypsin homology <TRY>
F,272,315,410/Active site: His, Asp, Ser #status predicted

Query Match 10.5%; Score 493; DB 2; Length 460;
Best Local Similarity 32.5%; Pred. No. 9.7e-25;
Matches 151; Conservative 47; Mismatches 196; Indels 70; Gaps 18;

```
QY 405 NGEKYCGERSQFV-----TSNKKITVRPH-SQSTTDGFLABLYSDSPCPQPTC 459
DB 45 NGQGYRGTSTTYTGKCKOSWSSMIPRHQKTPESYPNGLTMNYCRNPADKSPWCYT- 103
QY 460 RTGRCIRKE--LRCDGMDCTDSDELNCSGADHGFCKKPF-----CKPLFWVCD 510
DB 104 -TPPRVMEFCNLKAPQAPSVENPEPADCMIGIGYGRKATTVAGPQCE--WAAQE 160
QY 511 VNDGDNSEDOGSCPAQTFRCNGKCLSKSQOCCNGKDCGSDGSEASCPKNNVVTCTGH 570
DB 161 PRRHGIPTPE-----TNPRAGLEKVCYCRNPD--GD-----VNGPWCYTT 197
QY 571 TYRCLNGLCLSKNGPECDGKEDCSGSDSDE-KQDCGLRFTROARVGGTDADEGMPQ 629
DB 198 NPKRLDYC--DIPCESSFDGKRPKPKC-----PARVGGCVAPHPHMPQ 245
QY 630 VSLHALGQGHICGASLSPNVLVSAACHYIDDRGFRYSDPTQMTAFILGLHDQSORAPV 689
DB 246 VSLRRSRHFGCGTILSEWVLTANCLDSILG-----PSFTVLGHAYEAKRA-SV 289
QY 690 QERRRLRIISHPEFNDFTDYDIALLEKPAEYSSMVRPICLPDASHVFPAGKAIWTVG 749
DB 300 QEIPVGRFLPE-----SRADIALKLSSPAVITDEVIACPLSPBNYVADKTVCYTG 353
QY 750 WGHYVGGGALLIQGEIRVINQTCF--NLLPQITPRMNCVGLSGVSCQDSDG 807
DB 354 WGETQ--GTPGVGLKARLPVLENKVCNRYEYLNGVKSTELCAGLAGTDSQDSDG 412
QY 808 PLSSVEADGRIFGAGVSWGDCAGRNKPGVYTRLLPFRDWIKE 851
DB 413 PLVCFEKDKTIL-QSVTSMGLGCRPNKPGVYRVSTVYPIWE 455
```

RESULT 15

A29154
Complement factor I (EC 3.4.21.45) precursor - human
N/Alternate names: C3b/C4b inactivator
C/Species: Homo sapiens (man)
C/Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 09-Jul-2004
C/Accession: A29154; A28434; S66420
R/Cattell, C.F.; Lyons, A.; Sim, R.B.; Day, A.J.; Harris, T.J.R.
Biochem. J. 242, 849-856, 1987
A/Title: Characterization of the primary amino acid sequence of human complement control
A/Reference number: A29154; PMID:87241401; PMID:2954545
A/Accession: A29154
A/Molecule type: mRNA
A/Residues: 1-583 <CAT>
A/Cross-references: UNIPROT:P05156; GB:Y00318
R/Goldberger, G.; Brune, G.A.P.; Rice, M.; Edge, M.D.; Kwiatkowski, D.J.
J. Biol. Chem. 262, 10065-10071, 1987

A/Molecule type: protein
A/Residues: 483-507, 'E', 509-604 <M13>
R/Robbins, K.C.; Bernabeu, P.; Arzadon, L.; Summaria, L.
J. Biol. Chem. 248, 1631-1633, 1973
A/Title: The primary structure of human plasminogen. II. The histidine loop of human plasminogen.
A/Reference number: A92125; MUID:73149246; PMID:4694729
A/Contents: annotation; active site
R/Groskopf, W.R.; Summaria, L.; Robbins, K.C.
J. Biol. Chem. 244, 3590-3597, 1969
A/Title: Studies on the active center of human plasmin. Partial amino acid sequence of human plasminogen.
A/Reference number: A92048; MUID:6923373; PMID:4240117
A/Contents: annotation; active site
R/Texler, M.; Vail, Z.; Paddy, L.
J. Biol. Chem. 257, 7401-7406, 1982
A/Title: Structure of the omega-aminocarboxylic acid-binding sites of human plasminogen.
A/Reference number: A92382; MUID:82213905; PMID:6919539
A/Contents: annotation; omega-aminocarboxylic acid binding sites
R/Vall, Z.; Paddy, L.
Biochemistry 37, 13690-13694, 1998
A/Title: The fibrin-binding site of human plasminogen. Arginines 32 and 34 are essential for fibrin binding.
A/Reference number: A92458; MUID:85054794; PMID:6094526
A/Contents: annotation; fibrin binding site; omega-aminocarboxylic acid binding site
R/Go, Y.; Ji, R.W.; Davidson, D.; Schaller, J.; Marti, D.; Soehndel, S.; McCanne, S.G.; J. Biol. Chem. 271, 29461-29467, 1996
A/Title: Kringle domains of human angiotensin. Characterization of the anti-proliferative activity of the kringle domain.
A/Reference number: A58811; MUID:97067211; PMID:8910613
A/Contents: annotation
R/Lijnen, H.R.; Uggul, F.; Bini, A.; Collen, D.
Biochemistry 37, 4699-4702, 1998
A/Title: Generation of an angiotensin-like fragment from plasminogen by streptolysin-1 (M13).
A/Reference number: A58812; MUID:9548733; PMID:9548733
A/Contents: annotation
R/Mulichak, A.M.
submitted to the Brookhaven Protein Data Bank, July 1991
A/Reference number: A51341; PDB:1PK4
A/Contents: annotation; X-ray crystallography, 1.9 angstroms, residues 376-454
R/Tullinsky, A.; Wu, T.P.
submitted to the Brookhaven Protein Data Bank, July 1991
A/Reference number: A51488; PDB:2PK4
A/Contents: annotation; X-ray crystallography, 2.25 angstroms, residues 375-454
R/Wu, T.P.; Tullinsky, A.
submitted to the Brookhaven Protein Data Bank, August 1993
A/Reference number: A51911; PDB:1PKR
A/Contents: annotation; X-ray crystallography, 2.48 angstroms, residues 102-181
R/Padmanabhan, K.; Tullinsky, A.
submitted to the Brookhaven Protein Data Bank, April 1994
A/Reference number: A52408; PDB:1PMK
A/Contents: annotation; X-ray crystallography, 2.25 angstroms, residues 377-454
R/Tullinsky, A.; Mathews, I.I.
submitted to the Brookhaven Protein Data Bank, December 1995
A/Reference number: A65245; PDB:1CEB
A/Contents: annotation; X-ray crystallography, 2.1 angstroms, residues 102-181
R/Mulichak, A.M.; Tullinsky, A.; Ravichandran, K.G.
Biochemistry 30, 10576-10588, 1991
A/Title: Crystal and molecular structure of human plasminogen kringle 4 refined at 1.9 Å.
A/Reference number: A58819; MUID:92031502; PMID:1657148
A/Contents: annotation
R/Wu, T.P.; Padmanabhan, K.; Tullinsky, A.; Mulichak, A.M.
Biochemistry 30, 10589-10594, 1991
A/Title: The refined structure of the epsilon-aminocaproic acid complex of human plasminogen.
A/Reference number: A58818; MUID:92031503; PMID:1657149
A/Contents: annotation
R/De Vos, A.M.; Ulteich, M.H.; Kelley, R.F.; Padmanabhan, K.; Tullinsky, A.; Westbrock, M.; Biochemistry 31, 270-279, 1992
A/Title: Crystal structure of the kringle 2 domain of tissue plasminogen activator at 2.1 Å.
A/Reference number: A39483; MUID:92118803; PMID:1310033
A/Contents: annotation; X-ray crystallography, 2.4 angstroms
R/Sec, B.; Teeter, M.M.; Whitlow, M.; Yamano, A.
submitted to the Brookhaven Protein Data Bank, June 1995

A/Reference number: A65980; PDB:1KRN
A/Contents: annotation; X-ray crystallography, 1.67 angstroms, residues 376-454
R/Rejz, M.; Llinas, M.
submitted to the Brookhaven Protein Data Bank, August 1996
A/Reference number: A65803; PDB:1HPJ
A/Contents: annotation; conformation by (1)H-NMR, residues 103-181
R/Rejz, M.; Llinas, M.
submitted to the Brookhaven Protein Data Bank, August 1996
A/Reference number: A65804; PDB:1HPK
A/Contents: annotation; conformation by (1)H-NMR, residues 103-181
R/Rejz, M.; Llinas, M.
Eur. J. Biochem. 221, 927-937, 1994
A/Title: (1)H-NMR assignments and secondary structure of human plasminogen kringle 1.
A/Reference number: A65817; MUID:94237157; PMID:8181475
A/Contents: annotation; conformation by (1)H-NMR, residues 96-184
R/Rejz, M.; Llinas, M.
Eur. J. Biochem. 221, 939-949, 1994
A/Title: Solution structure of the epsilon-aminocaproic acid complex of human plasminogen.
A/Reference number: A58817; MUID:94237158; PMID:8181476
A/Contents: annotation; conformation by (1)H-NMR
C/Comment: Plasminogen is synthesized by the kidney and is present in plasma and many other tissues. Plasminogen is converted to plasmin by plasminogen activators (see PIR:UKHU and PIR:GHCUB).
C/Comment: Plasmin is inactivated by alpha-2-antiplasmin (see PIR:ITHUA2) immediately after release from the fibrinolytic system. Plasminogen is formed by autolytic cleavage of plasmin under artificial conditions.
C/Comment: Streptolysin 1 (see PIR:KCHUS1) acts on plasminogen to produce angiotensin. Tissue plasminogen activator.
C/Genetics: GDB:GDB:GDB
A/Genes: GDB:GDB:GDB
A/Cross-references: GDB:119498; OMIM:173350
A/Map position: 6q26-6q27
A/Intons: 17/1; 62/2; 98/1; 136/2; 183/1; 223/2; 263/1; 317/2; 366/1; 419/2; 480/1; 529/1
A/Function: A/Description: dissolves the fibrin of blood clots; acts as a proteolytic factor in various tissues; also activates the urokinase-type plasminogen activator.
A/Pathway: fibrinolysis
C/Superfamily: plasmin; kringle homology; plasminogen-related protein precursor homology; C/Keywords: angiotensin inhibitor; blood; duplication; fibrinolysis; glycoprotein; hydrolase; F1-19/Domains: plasminogen-related protein precursor homology <PLPH>
F1-19/Domains: signal sequence; status predicted <SIG>
F20-810/Product: plasminogen #status experimental <PRO>
F20-96/Domains: activation peptide #status experimental <APR>
F79-466/Product: angiotensin #status experimental <AST>
F79-580/Domains: plasmin chain A #status experimental <CHA>
F103-181/Domains: kringle homology <KR1>
F185-262/Domains: kringle homology <KR2>
F275-352/Domains: kringle homology <KR3>
F377-454/Domains: kringle homology <KR4>
F461-560/Domains: kringle homology <KR5>
F550-580,581-810/Product: microplasmin #status experimental <MWT>

Query Match 10.4%; Score 488; DB 1; Length 810;
Best Local Similarity 24.5%; Pred. No. 4e-24;
Matches 223; Conservative 103; Mismatches 296; Indels 288; Gaps 49;
DB 22 LDVYVNTQGSLSVTKQ-----LGASIECAKCED-----EEFTC 61
QY 102 VDAVENSSTEFSLASKYVDALKLYSGVFPGLPYHKSAYTAFSGSVIAYWSEF--159
DB 22 LDVYVNTQGSLSVTKQ-----LGASIECAKCED-----EEFTC 61
QY 160 -SIPQHLVEAEERVAEERVMLEPPRAISLKSFTVTVAAFPDTSK-----TVQRT 209
DB 62 RARQVYSKEQGVYMAENRKSITIRWDRV---VLEPKKYVLSCEKTKGNGKRYGTMSKT 118
QY 210 QDNSSCSFGHLARVEMLRFT-----PGPDPSPYRNA-----RCQ-----MALRGP 251
DB 119 KN-----GITQKWSSTSPHPRSPAPHPBEGLEBNYCRNDPDGQPPWCTYTD 168
QY 252 ADVSLTSPSPFLASCDE-----GSDU-VTVYVNTLSPME-----PHALVOLCGTY 297
DB 169 PER-----RIDYCDILCEBCECHGSGENYDGIKSKTMSGLGCAQMSQSFHA-----HGVI 220

Qy	298	P--PSNLTHSSONVLLTLTNTERR-----HPFEATPOLPMS-----SCGRL	344
Db	221	PSKRPNNKLKKNYGRN-----PDRELRFPCFTTDPKRNKELCDIRCTTPRSSSGPT	273
Qy	345	RKAQGTNSPYYPGHYPNNIDCT-----WNIEVPNNQHVKRFKPYLLBEGVAGT	396
Db	274	QCLGKTGE-----NYRGNVALVTSGHTCQHMSAQTPHTN-----R	309
Qy	397	CPKXYVEIN-GEKVC-----GERSQFVTSNSNKLTVPNHSQSYTDIDGFLAELYSYSD	451
Db	310	TENPFPCKNLDENYCRNPDGKRAAPMCHTTSQ-----VNR-----EYKIKISCD	353
Qy	452	PCP-----GQ-----FTCRGRCIRKELRCDMADTDH	480
Db	354	SSPVTBQLAPLAPPELTPVVDQCYHGDQGYRGTSYTTTGK-----KCSQSMSTPH	407
Qy	481	SDELNCGSDAHQOTCKKKKPKRPLFWCDSDVNDGSDSDEGSCS-----PAQTR-CSNG	535
Db	408	R-----HQKTPENYPNAGL-----TNNYCRNPDADKGPWCFTTDPBSVRENYCNLK	452
Qy	536	KC-----LSKSQCCNGKDC- GDG-----SDBAS-----	558
Db	453	KCSGTASVAPRPVVLLENVETPSBEDCMGNKGVRKRAATVVTGTPQCDMAAQEPNR	512
Qy	559	----CPKXNVVTCIKHTYRCLN--GLCLSKNPEC--DKEKDCSDGSDERDC-----DCG	605
Db	513	HSITFPTNPRAGLEKUY-CENPDG--DVGSPCYTTPNPKLYDCVDPQCAPSPDCG	568
Qy	606	LRSTFRQ---ARVVGSTDADEGEMPMOVSLLALQGNHCGASLISPMVLVAACHYIDR	662
Db	569	KPQVEPKKCPKRRVVGCVAHNPSMPMOVSILTRGCMFCGSLISPEWVLNACL--EK	626
Qy	663	GFRYSDEPTQWTAFLGLHDQSORSAPOVERLKKIISHPFNDFTFYDIALLELEKPAE	722
Db	627	SPR--PSSYKVIILGAH-QEVNLBPHVOEIVSRILFLEPTRK-----DIALTLKISSPAV	676
Qy	723	YSSWVRPLCLBDASHVFPAGKAIWVTGHTGYOYSTGALILQKEIIVYNTQTE--NL	780
Db	677	ITDKVITPACLSPPNYVADRTCECTITGGEYO-GLFGGLKEKQOLPYIEKKVCNRRBEFL	735
Qy	781	POQITPRMVCVPLSGVDSQQSDSGGRLSSVEADGRIFGAGVYSWGDGCAGRNKPQVYT	840
Db	736	NGRVQSTELCAGHLAGTDSQQDSDGGPLVCFEKDKYIL-QGVYSWIGSGCARPKRPGVVY	794
Qy	841	RLPLFRDMIX 850	
Db	795	RVSRRFVTWIE 804	

RESULT 17
KEBO
coagulation factor IXa (EC 3.4.21.22) precursor - bovine
N/Alternate names: Christmas factor
C/Species: Bos primigenius taurus (cattle)
C/Date: 30-Nov-1980 #sequence revision 03-Aug-1984 #text_change 09-Jul-2004
C/Accession: A14757; B20274; I45891; A00923
R/Katayama, K.; Ericsson, L.H.; Entfield, D.L.; Walsh, K.A.; Neurath, H.; Davie, E.W.; Thrombolytic
A/Title: Comparision of amino acid sequence of bovine coagulation factor IX (Christmas factor) and human factor IX (Christmas factor)
A/Accession: A14757
A/Reference number: A14757; MUID:80056619; PMID:291916
A/Molecule type: protein
A/Residues: 1-63, 'T', 65-416 <KAT>
A/Cross-references: UNIPROT:P00741
R/McMullen, B.A.; Fujikawa, K.; Kistiel, W.
Biochem. Biophys. Res. Commun. 115, 8-14, 1983
A/Title: The occurrence of beta-hydroxyaspartic acid in the vitamin K-dependent blood coagulation factor IX
A/Reference number: A20274; MUID:83308813; PMID:6688526
A/Accession: B20274
A/Molecule type: protein
A/Residues: 59-63, 'X', 65-69 <MCN>
R/Choo, K.H.; Gould, K.G.; Rees, D.J.G.; Brownlee, G.G.
Nature 299, 178-180, 1982

A>Title: Molecular cloning of the gene for human anti-haemophilic factor IX.
A>Reference number:145891; MUID:82272386; PMID:6287289
A>Accession: 145891
A>Status: translated from GB/EMBL/DBJ
A>Molecule type: mRNA
A>Residues: 52-139 <CHO>
A>Cross-references: GB:J00007; NID:9163053; PIDN:AAA0520.1; PID:g163054
J>Hase, S.; Kawabata, S.; Nishimura, H.; Takeya, H.; Sueyoshi, T.; Miyata, T.; Iwanaga, S.
J>Biochem. 104, 867-868, 1988
A>Title: A new trisaccharide sugar chain linked to a serine residue in bovine blood coagulation factor IX.
A>Reference number: A4456; MUID:89213999; PMID:3149637
A>Contents: annotation
A>Note: structure and location of a carbohydrate covalently bound to Ser
C>Comment: Factor IX is activated by factor XIa, which excises the activation peptide prior to the activation of factor X.
C>Comment: The gamma-carboxyglutamic acid residues arise by posttranslational, vitamin K dependent modification of the gamma-carboxyglutamic acid (Gla) residues and, with stroke, are essential for the biological activity of the protein.
C>Function:
A>Description: catalyzes the proteolytic activation of coagulation factor X in the presence of factor V and phospholipids.
A>Pathway: blood coagulation intrinsic pathway
C>Superfamily: coagulation factor X; EGF homology; Gla domain homology; trypsin homology
C>Keywords: beta-hydroxyaspartic acid; blood coagulation; calcium binding; carboxyglutamic acid
F1-146/Product: coagulation factor IXa light chain #status experimental <ALC>
F1-145/Domain: Gla domain homology (fragment) <GLA>
F1-85/Domain: EGF homology <EG1>
F1-81-82/Domain: EGF homology <EG2>
F1-81-124/Domain: EGF homology <EG3>
F1-147-181/Domain: activation peptide #status experimental <APT>
F1-181-416/Product: coagulation factor IXa heavy chain #status experimental <AHC>
F1-182-409/Domain: trypsin homology <TRY>
F1-78 15 17 20 21 26 27 30 33 35 40/Modified site: gamma-carboxyglutamic acid (Glu) #status experimental
F1-18 23 51 62 56 71 73 82 88 99 95 109 111 124 133 230 207 223 337 351 362 390/Disulfide
F153/Binding site: carboxylate (Ser) (covalent) #status experimental
F164/Modified site: erythro-beta-hydroxyaspartic acid (Asp) #status experimental
F165 168 173 261/Binding site: carboxylate (Asn) (covalent) #status experimental
F1222 270 366/Active site: His, Asp, Ser #status predicted

```

OY 399 KDYAEINEKCYGSESOQFVWTSNSNKTIVRFHSDQSYTGTGFLAEYLSYSOSDPGCOFT 458
Db 7 BEFVRGNLERBEKEK-----CSFEAREVFENTEXTTE--FKQYVDGQ----- 50
OY 459 CRTGRCIRKEKLRCDGWADCTDHSDBLSCDAGHOFT-----CANKFCKPL 504
Db 51 CESNPCLNGM-----CKDDINSYECWQAGFEBCNCELDATCSIKNRCCK-QOFCK-- 100
OY 505 FHWCDSDVNDGDNSDGQSCPAQFFRCGNGKCLSKSQOCCNGKDDCGDGSDEASCPRYNV 564
Db 101 -----RTDKNKVVCS-----CTDGRRLADQK-----SC-BPAVPPCCGAVSV 137
OY 565 VTCTGHTYRCINGCLCLSKGNPECDGK-EDCSGSDSEKDDCDGRLSFTROARVVGSTDADE 623
Db 138 SHISKKLURA-ETIFSMNTYENSSEALITMWNVTQSN-----QSFDEFGRVVGGEBAER 190
OY 624 GEMPRQVSLHALGQGT---CGASLISPMWLYSAACYIDDRGFRISDPQMTAFGLHD 680
Db 191 GQFPMQVILH-----GBIAFCGGSIVNERKVVYTAAC-----IKPGVKTIVVAEBHN 238
OY 681 OSORAPGVOEBRLKRIISHPPFNDF--TFVDIALLLEKPAEYSSMVRPICLPDA--S 736
Db 229 -TEKEPFEQKKNVRAIRALPYHYNASINKYSHDIALLEDEPELJNSYVTPICLADDTY 297
OY 737 HVFPAGKAIWYTWGHTOYGTGALILQKGEIRVINQTTCENILPOOIITPRMNCVFLSG 796
Db 298 NIFSFGYGVYSGMWKVFNRGRSASILQYLAKPLVDARCLSTKTSIYSHMCAGYHEG 357
OY 797 GVDSCGSGGGLSVSEADGRIFGAGVYVMGGCAGARNRPGVYTRILPFDWIKENT 853
Db 358 GKDSCQGSQSGG- HYTEVEGTSFLTGILISWGECAKNGKGYLITKSSRYNNMIKEKT 413

```

RESULT 18

QY 351 ----FNSPYFGHYPNIDCTMNIEV-----PNNQ-----HYKRFKFEVLE----- 389
 Db 304 QRSBQTPHKNRPPENPPCK-NLEBNYCRNPNGEKAWCYTTSSEVWEICTIPSCSS 362
 QY 390 -----PGVPAGT-CPKDYVEINGEKYCGERSQFVV-----TSNSKITVRFH--SDQS 434
 Db 363 PLSTERMDVPVBPBQTPVPQDCYHNGQSYNGSTSTTITGRKCSQWSMTDHRHLKTPEN 422
 QY 435 YTDGFLAEYLSTYSSSPDCPGQFT-----CRGECIR--KELRCDGADCTDHSDE 483
 Db 423 YPNAGLTMMNYCRNDADKSPWCYTTPDPRVWEFCNLKKCSBTPPPAPAPQAVENPE 482
 QY 484 LNCSCDAGHOFTCKNKF-----CKPLFWVCDNVNDGDNDEQSCCPAQTPRCNSGKC 537
 Db 483 ADCNIGTKSKYRGKATTVAGVPCOE--MAAOEPHSHIFRPE-----TNFQS 528
 QY 538 LSKSQCGKXKDCDGDGSDDEACPKRVNVYCTKHTYRCLNGCLSKNGECCDKEDCDGSG 597
 Db 529 GLEBNYCRNPD--GD-----VNGHWCTMNPBRKPFYD---DVPQCSSTPDCGKPK 574
 QY 598 DE-KDCDCGLRSFTROARVVGCTDADGEMWQVSHALGGHICGASLISPMVLVSAH 656
 Db 575 VEPKCC-----SGRIVGCVSKPHSWPMQVSLRSSR--HFCGGLTISPKNVLTAMH 624
 QY 657 CYIDRGFRVSDPTQWTAFLGLHDQSORAPGOERRLKRIISHPFNDFTFDYDIALLE 716
 Db 625 CLNDITLALSF-----YKILGANHEKYTE-OSVDEIPVSRILFRPD-----SQADIALLK 672
 QY 717 LEKPAEYSSMVRPILCLPDASHVPFAGKAIWYTGWHTQYGGTGAALLQKGEIRVYNQTT 776
 Db 673 LSRPAITTKVEIYIPACLPFPNVMVARTCYITGWEQT-GTFGGGLLKEALPIYENKVC 731
 QY 777 --ENLLPQOITPRMCMCVFLSGVDSCGDSGGLSSVADGRIFGAGVWVGCGCAGR 834
 Db 732 NRNEYLLGRVVPTEICGHLIGTDSGCGSGPLVCPKDKYTL-QGVTSWGLGCARP 790
 QY 835 KPGVYTRLPFRDWIKE 851
 Db 791 KPGVYTRVSPVPIEE 807

RESULT 20

A57014
 prostaasin (EC 3.4.21.-) precursor - human
 C/Species: Homo sapiens (man)
 C/Date: 24-May-1996 #sequence_revision 24-May-1996 #text_change 09-Jul-2004
 C/Accession: A57014; A54866
 R/Yu, J.X.; Chao, L.; Chao, J.
 J. Biol. Chem. 270, 13483-13489, 1995
 A/Title: Molecular cloning, tissue-specific expression, and cellular localization of human prostaasin
 A/Reference number: A57014; MUID:95286644; PMID:776952
 A/Accession: A57014
 A/Status: translated from GB/EMBL/DBJ
 A/Molecule type: mRNA
 A/Residues: 1-343 <RES>
 A/Cross-references: UNIPROT:Q16651; GB:LA1351; NID:9862304; PIDN:AA611759.1; PID:9862305
 A/Experimental source: prostate
 A/Note: parts of this sequence were determined by protein sequencing
 R/Yu, J.X.; Chao, L.; Chao, J.
 J. Biol. Chem. 269, 18843-18848, 1994
 A/Title: Prostaasin is a novel human serine proteinase from seminal fluid. Purification, characterization, and tissue distribution
 A/Reference number: A54866; MUID:94308140; PMID:8034638
 A/Accession: A54866
 A/Molecule type: protein
 A/Residues: 45-64 <YUN>
 C/Genetics:
 A/Gene: GDB:PRSS8
 A/Cross-references: GDB:676446; OMIM:600823
 A/Map position: 16p11.2-16p11.2
 C/Superfamily: trypsin; trypsin homology
 C/Keywords: glycoprotein; hydrolase; serine proteinase; transmembrane protein
 F/1-32/Domain: signal sequence #status predicted <SIG>
 F/33-44,45-343/Product: prostaasin #status predicted <MTR>

F/33-44/Domain: prostaasin light chain #status predicted <CHL>
 F/45-343/Domain: prostaasin heavy chain #status predicted <CHH>
 F/45-281/Domain: trypsin homology <TRY>
 F/323-341/Domain: transmembrane #status predicted <TMN1>
 F/37-154,70-86,168-244,201-223,234-262/Disulfide bonds: #status predicted
 F/85,134,238/Active site: His, Asp, Ser #status predicted
 F/159/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 10.3%; Score 483.5; DB 1; Length 343;
 Best Local Similarity 39.3%; Pred. No. 2.9e-24;
 Matches 105; Conservative 42; Mismatches 95; Indels 25; Gaps 8;
 QY 596 GSDKDCDGLRSFTROARVVGCTDADGEMWQVSHALGGHICGASLISPMVLVSA 655
 Db 29 GABEAEPQCV--APQARITGSSAVAGQPMQVSTTIGV--HYCGSLVSEKVVLSA 84
 QY 656 HCYIDRGFRVSDPTQWTAFLGLHDQSORAPGOERRLKRIISHPFNDFTFDYDIAL 715
 Db 85 HCFPSEH-----HKEAVEVKGAN--QDSYSEDAKVSTLKDIIHPFVYLQEGSGQDIAL 138
 QY 716 ELEKPAEYSSMVRPILCLPDASHVPFAGKAIWYTGWHTQYGGTGAALL--LQKGEIRV 771
 Db 139 QLSRPITFSRYIRPILCPANASFPNGLHCTVTGWHV--APSVSLTPKPLQQLVPLI 196
 QY 772 NQTCENML-----PQOITPRMCMCVFLSGVDSCGDSGGLSSVADGRIFGAGV 823
 Db 197 SRETCNCLNIDAKPEPFPVQEDMVCAGYVEGSKDCQDGGPL-SCVEGLMYLTGI 255
 QY 824 VSMGDCGAGRNKPGVYTRLPFRDWIK 850
 Db 256 VSMGDCAGARNRPGVYTRVLSVSWIQ 282

RESULT 21

PLMS
 Plasmin (EC 3.4.21.7) precursor - mouse
 N/Contains: angiotatin; plasminogen
 C/Species: Mus musculus (house mouse)
 C/Date: 20-Sep-1991 #sequence_revision 01-Nov-1996 #text_change 09-Jul-2004
 C/Accession: A38514; S48202; S48203
 R/Degen, S.J.F.; Bell, S.M.; Schaefer, L.A.; Elliott, R.W.
 Genomics 8, 49-61, 1990
 A/Title: Characterization of the cDNA coding for mouse plasminogen and localization of the gene
 A/Reference number: A38514; MUID:91164812; PMID:2081600
 A/Accession: A38514
 A/Molecule type: mRNA
 A/Residues: 1-812 <DEG>
 A/Cross-references: UNIPROT:P20918; GB:J04766; NID:9200402; PIDN:AAA50168.1; PID:9200403
 R/Lijnen, H.R.; van Hoef, B.; Beelen, V.; Collen, D.
 Eur. J. Biochem. 224, 863-871, 1994
 A/Title: Characterization of the murine plasma fibrinolytic system.
 A/Reference number: S48202; MUID:95010076; PMID:7523120
 A/Accession: S48202
 A/Molecule type: protein
 A/Residues: 20-25 <LIJ>
 A/Accession: S48203
 A/Molecule type: protein
 A/Residues: 22-27 <LIJ>
 C/Comment: Plasminogen is synthesized by the kidney and is present in plasma and many other tissues. Plasminogen is converted into plasmin by plasminogen activators, both plasminogen activator and plasminogen activator. In the presence of the inhibitor, the activation of plasminogen involves also removal of the activation peptide.
 C/Comment: Streptokinase (see FIR:KCMSST) acts on plasminogen to produce angiotatin. The enzyme in treating solid tumors.
 C/Function:
 A/Description: dissolves the fibrin of blood clots; acts as a proteolytic factor in a variety of the walls of the graafian follicle; also activates the urokinase-type plasminogen activator
 A/Pathway: fibrinolysis
 C/Superfamily: plasmin; kringle homology; plasminogen-related protein precursor homology;
 C/Keywords: angiogenesis inhibitor; blood; duplication; fibrinolysis; glycoprotein; hydrolase
 F/1-96/Domain: signal sequence #status predicted <SIG>
 F/1-19/Domain: signal sequence #status predicted <SIG>
 F/20-812/Product: plasminogen #status predicted <PRO>

F;20-96/Domain: activation peptide #status predicted <APr>
 F;79-466/Product: angiotensin #status predicted <AST>
 F;97-581,582-912/Product: plasmin #status predicted <AMT>
 F;97-581/Domain: chain A #status predicted <ACH>
 F;103-181/Domain: kringling homology <KR1>
 F;185-262/Domain: kringling homology <KR2>
 F;275-352/Domain: kringling homology <KR3>
 F;377-454/Domain: kringling homology <KR4>
 F;481-560/Domain: kringling homology <KR5>
 F;582-812/Domain: chain B #status predicted <BCH>
 F;582-805/Domain: trypsin homology <TRY>
 F;49-73,53-61,103-181,124-164,152-176,185-262,188-316,206-245,234-257,275-352,296-335,32
 Bonds: #status predicted
 F;78-79/Cleavage site: Glu-Asn (stromelysin 1) #status predicted
 F;136,308/Cleavage site: carboxylate (Asn) (covalent) #status predicted
 F;466-467/Cleavage site: Thr-Val (stromelysin 1) #status predicted
 F;581-582/Cleavage site: Arg-Val (plasminogen activator) #status experimental
 F;624,667,762/Active site: His, Asp, Ser #status predicted

Query Match 10.3%; Score 483.5; DB 1; Length 812;
 Best Local Similarity 26.1%; Pred. No. 7.8e-24;
 Matches 189; Conservative 76; Mismatches 277; Indels 181; Gaps 34;

Qy 228 FTTPGPDSPY-----PAHRCQMALRGADSVLSITFRSFDLASCDEKSDLVVYNT 281
 Db 165 YTTD--FDKRYDYCNIECEECMYCSGEKYEKISKTMGLDQAWDSQ----- 212
 Qy 282 LSPMEPHALVOLCTYP--PSYVLTFRSSONVLLITLITNTER-----HGFBAATF 332
 Db 213 -----SPHA-----HCYIAKPPSKLKNYCHN-----PDGERPMCFITTDPKREYC 257
 Qy 333 QLPMSM-----CGRLKKAQGTN-----SPYRPHYPPNIDCT 367
 Db 258 DIRCTTPPPSPPTVQCLKRGGENYRGTVSVVSGKQWBSQTHRHNRTPENPC 317
 Qy 368 WNE-----VNNQHVAKRFKFFILBEGVAGCP-----KD 400
 Db 318 -NIEENYCRNPDGTAWCYTTDSQLMEYCEIIPSCSSASPDGSDSVPEEGTPVQ 376
 Qy 401 YVEINKECYGERSQFVNTSNKITVRF--HS--DOSYDTGFLAELYSYDSDCPG 455
 Db 377 CYGSDGSGYKGTSTTTTGKCCGMAAMFPHRSKTEBNFPDGLNNYCRNPDGSPW 436
 Qy 456 QFTCRIGRCIRKELKCDGMDCTDSDLELNCSDAGHQFTCKKFKPLFWVCDVND 514
 Db 437 CYT--TDPVYR-----MEYC-----NLKRCSETGSGVLELPTVSGP--SGPSDSERDCM 482
 Qy 515 -GNSDSEQ-----GCPAQTFC-----SNCKLSKSGQCCNGKDCGSD 555
 Db 483 YGNGKDYRGKTAATACTPCQGMADPHRHSIFTPTNPRADLEKVCNPD--GD-- 537
 Qy 556 EASCPKVNVTCTGTYR-----CLNGCLSKNPECDGKEDSDGDEKDCDGLRFT 610
 Db 538 -----VNGPWCYTTNPKLYDYDIPLCASASFPCC--GKPV-----EPKCC----- 578
 Qy 611 RQARVVGCTDADEGEPMQVSLHALGQG--HI CGASLISPMWLSAAHCYIDRGRFRYSD 669
 Db 579 -PGVVGCGCAANPSPWQISLIRFTGQHFCCGGLTAPMVLTAACHL-----EKSSRP 632
 Qy 670 TQMTAFGLGDSQSRSAFGVQERLKRILSHFPNDTPFDVLLALEKPAEYSSWRP 729
 Db 633 EFKVILIGAHEEYIRGL--DVQELISVALILBP-----NNRDIALKLSPATTTDVIIP 685
 Qy 730 ICLPDASHVPAGKAIWVTGHTGYGCTGALLLQKEIRVINGTCEM--LPQQTTPR 787
 Db 686 ACLEPSPYMWADRICTITGMEGTQ--GTFPAGRLKELQVLEIKVCNRYEILNNRKST 744
 Qy 788 MNCVGLISGVDSQCGDSSGSPPLSSVEADGRIFGAGVSVWSDGCGARKKPGVYTRLLPLFRD 847
 Db 745 ELCAQGLAGVDSQCGDSSGSPPLVCFEKKYIL--QGVTSWGLGCGARPKKPGVYVVSFVD 803
 Qy 848 WIK 850
 ||:

Db 804 WIE 806

RESULT 22

353777
 hepsin (EC 3.4.21.-) - rat

C;Species: Rattus norvegicus (Norway rat)

C;Date: 06-Jan-1995 #sequence_revission 06-Jan-1995 #text_change 16-Aug-2004

C;Accession: S33777; S32013

R;Farley, D.; Raymond, F.; Nick, H.

Biochim. Biophys. Acta 1173, 350-352, 1993

A;Title: Cloning and sequence analysis of rat hepsin, a cell surface serine proteinase.

A;Reference number: S33777; PMID:93305733; PMID:8318546

A;Residues: 1-416 <RNA>

A;Cross-references: UNIPROT:Q05511; EMBL:X70900; NID:G57928; PIDN:CAA50256.1; PID:G57929

C;Superfamily: trypsin homology

C;Keywords: hydrolase; liver; serine proteinase; transmembrane protein

F;22-44/Domain: transmembrane #status predicted <TMN>

F;162-399/Domain: trypsin homology <TRY>

F;187-203,290-358,321-337,348-380/Diulfide bonds: #status predicted

F;202,256,352/Active site: His, Asp, Ser #status predicted

Query Match 10.2%; Score 479.5; DB 1; Length 416;
 Best Local Similarity 38.9%; Pred. No. 6.6e-24;
 Matches 107; Conservative 42; Mismatches 91; Indels 35; Gaps 11;

Qy 603 DCLRSTQARVVGCTDADEGEPMQVSLHALGQGHICGASLISNWLVSAAHCYIDDR 662
 Db 151 DCRRLKLPVD--RIVGGQSDSLGRMPQVSLRYDGT--HLGGSLLSDWVLTAAHCPPE-- 206

Qy 663 GFRYSDPTQWTAFLGLHDSQSRAPGQERLKRISH-----PFNDFPD--YDIAL 715
 Db 207 --RNRVLSRKRVAG--AVARTSPHVAQVGVAVLYHGILP--FDDPTIDENSDIALV 260

Qy 716 ELKPEAYSSWVRICLPDASHVFPAGKAIWVTGHTGYGCTGALLQKEIRVINGT 775
 Db 261 HLSSSLPFLFYICLPVGLPAAGALVNDGKCTVGMGNTGYGQAAVVLDEARVILISNEV 320

Qy 776 CE--NLLPQITRMNCVGFSLSGVDSQCGDSSGSPPLSSVEADDRIFG-----AGVSWG 827
 Db 321 CNSPDYGNQIKKMKCAGPREGIDACQGDSDG--HFVCEDRISGTSRMLCGIVSWG 377

Qy 828 DGCAGNKKGVYTRLLPLFRDWI-----KENTGV 855
 Db 378 TGCAARKGVYTKVIVDFRHWIRQALKTHSSEATGM 412

RESULT 23

B30848
 plasmin (EC 3.4.21.7) precursor - rhesus macaque

C;Species: Macaca mulatta (rhesus macaque)

C;Date: 31-Mar-1989 #sequence_revission 31-Mar-1989 #text_change 09-Jul-2004

C;Accession: B32869; B30848

R;Tomlinson, J.E.; McLean, J.W.; Lawn, R.M.

J. Biol. Chem. 264, 5957-5965, 1989

A;Title: Rhesus monkey apolipoprotein (a). Sequence, evolution, and sites of synthesis.

A;Reference number: A32869; PMID:89174660; PMID:2525643

A;Accession: B32869

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-810 <TMN>

A;Cross-references: UNIPROT:P12545; GB:J04697; NID:G342272; PIDN:AAA36901.1; PID:G342273

C;Superfamily: plasmin; kringling homology; plasminogen-related protein precursor homology;

C;Keywords: fibrinolysis; glycoprotein; hydrolase; kringling; serine proteinase

F;1-9/Domain: signal sequence #status predicted <SIG>

F;103-181/Domain: kringling homology <KR1>

F;185-262/Domain: kringling homology <KR2>

F;275-352/Domain: kringling homology <KR3>

F;377-454/Domain: kringling homology <KR4>

F:481-560/Domain: kringles homology <KR5>
 F:581-803/Domain: tryptsin homology <TRY>
 F:49-73, 53-61, 103-181, 124-164, 152-176, 185-262, 188-316, 206-245, 234-257, 275-352, 296-335, 332
 Bonds: #status predicted
 F:622, 665, 760/Active site: His, Asp, Ser #status predicted

Query Match 10.2%; Score 477; DB 2; Length 810;
 Best Local Similarity 24.5%; Pred. No. 2, 1e-23;
 Matches 226; Conservative 96; Mismatches 288; Indels 312; Gaps 50;

QY 102 VDVAENSNSTFVSLASKVKDALKLYSGVPLFGYHKEASAVTAFSEGSVIAVYWSF- 159
 DB 22 LDDVYNTGASLFFITKKQAGSI-----SECAKCEEE-----EEFTC 61
 QY 160 -SIQHLVEAEARMAERVMVLPRAASLKSFVTVSVAP----- 200
 DB 62 RSFGYHSKEQOCVMAENR-----KSIVFRNRDVLEKKVYISECTGNGKNYRG 113
 QY 201 TDSKTVOITQDNSSCFGLHARGVELMRFT-----PGPPDSPYPAHA-----RCQ----- 245
 DB 114 TMSKT--RT-----GITCQKWSSTSPHRTSPATHSEGLSENYCENPDNDG 159
 QY 246 ---WALRGDADSVSLTFRSPDLASCDER-----GSDL-VTVYNTLSPEM-----PH 288
 DB 160 QGPMCYTTDPRE---REDYCDIPECEDECHMGSENYDGKISKTMSGLEQAMDQSOPH 215
 QY 289 ALVOLCGYTP---PSYNLTFHSSQNVLLITLITRER-----HGEFAFFQLPMS- 338
 DB 216 A---HGITPEKFPNKILKKNYCN-----PGDEPPMCFITDPNKMELCDI PRCTT 264
 QY 339 ---SCGRLRAQGTFSNPPYGHYPPNID--C-TWNIENVNNOHVAREKFFYLEPG 391
 DB 265 PPSSGPTVQCLKKTGEN--YRGDAVAVVSGHTCHGSAQPHHN----- 308
 QY 392 VPACTCPDYVEIN-GEKYC---GERSQFVYTSNKNITVR----- 428
 DB 309 ---RTENFPCKULDENYCNRPGEKAPMCTYNSQ---VRMEYCKIPSCSSPVSTEP 361
 QY 429 -----FHSD-QSYTDGFLAEVLSYSDSDCPQOFCRTCRCKEIRC 471
 DB 362 LDPAPPELTPVQVECHYGQDSYRG-----SSTTTTK-----KC 398
 QY 472 DGMADCTHSDDELNCSCDAGHQFTCKNFKCPFLFVVCDSVNDGDNDSQSC---PA 527
 DB 399 QSNWSMTPH-----WHKTPENPNAGL-----TMYCNPNADKPKMCFITDPS 443
 QY 528 QTFRCNSGKCLSKQOCNG-----KDC--GPDSEASCPRVNV 564
 DB 444 VRMEYCNLK-----KCSGTGSVAAPPPVAQLPDAETPSBEDCMFGNGKGRG-KAAT 496
 QY 565 VTCT-----KHYRCLN-----GL-----CLSK-----GNPEC--DKEKDCSDSD 598
 DB 497 VTGPRCEMAAQBESHRIFTPETNPRALEKNYCRNPDGDVGPMCTTNPRLKLFYCD 556
 QY 599 EKDC-----DGLRSFTRO---ARVGGTDADEGEPMQVSLHAGQHLGASLSPNV 650
 DB 557 VPQDASSFDGCKQVKKCPRGVVGCAVAFHSMWQIISLRLRGLHGFSGILLBEM 616
 QY 651 LVSAAHCYIDRGRFYSDPTQMTAFGLHDSQSBAPQVORRLKRIISHFENDFTDY 710
 DB 617 VLTAAHCL-----EKSSRPSPFYKYLGAH-REVLHPVQIEIVSKMSEP-----ARA 664
 QY 711 DIALLEKPAEYSWRFICLPDASHVFPAGKAIWVGWHTQVGGIALLQGEIRV 770
 DB 665 DIALKISSPAIITDKVLPACLPSPNYVADRETCFITGMGETY-GTYGAGLLKEARLPV 723
 QY 771 INQTTCE--NLLPQOITPRMVCVFLSGVDSQGDGSPISYEADGRIRGAGVWSGD 828
 DB 724 IENKVCNRYEFLNLTGVTTELCAHLAGTDSQGDSDGRLVCFBKXKYL--QGVTSKGL 782
 QY 829 GCAGRNKRGVYTRLLPRDWIK 850
 DB 783 GCARNKRGVYTRVRSFVTWIE 804

RESULT 24

A59271
 Ra-reactive factor (EC 3.4.21.-) 2 precursor - human
 N:Alternate names: mannosyl binding protein-associated serine proteinase 2 (MASP-2)
 C:Species: Homo sapiens (man)
 C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 09-Jul-2004
 C:Accession: A59271
 R:Title: S.; Vorup-Jensen, T.; Stover, C.M.; Schwaeble, W.J.; Laureen, S.B.; Poulsen, K.
 Nature 386, 506-510, 1997
 A:Title: A second serine protease associated with mannan-binding lectin that activates C
 A:Reference number: A59271; PMID:97242412; PMID:9087411
 A:Accession: A59271
 A:Status: nucleic acid sequence not shown; not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-686 <JEN>
 A:Cross-references: UNIPROT:O00187; GB:Y09926; NID:G4007626; PIDN:CAA71059.1; PID:G40076;
 A:Experimental source: tissue liver
 A:Note: submitted to GenBank, December 1996
 A:Note: parts of this sequence, including the amino end of the mature protein, were deter
 C:Genetics:
 A:Gene: GDB:MASP2
 A:Cross-references: GDB:6071500
 A:Map position: 1p36.2-1p36.3
 C:Superfamily: complement-activating serine proteases C1r/C1s/MASP; C1r/C1s repeat homol
 C:Keywords: beta-hydroxyasparagine; complement pathway; duplication; hydrolase; serine pr
 F:1-15/Domain: signal sequence #status predicted <SIG>
 F:16-44, 445-686/Product: Ra-reactive factor 2 #status predicted <MAT>
 F:19-134/Domain: C1r/C1s repeat homology <C1R1>
 F:142-180/Domain: EGF homology <EGF>
 F:184-293/Domain: C1r/C1s repeat homology <C1R2>
 F:300-361/Domain: complement factor H repeat homology <FH1>
 F:366-430/Domain: complement factor H repeat homology <FH2>
 F:445-679/Domain: tryptsin homology <TRY>
 F:72-90, 142-156, 152-165, 167-180, 184-211, 241-259, 300-348, 328-361, 366-412, 396-430, 434-552,;
 F:158/Modified site: erythro-beta-hydroxyasparagine (asn) #status predicted
 F:444-445/Cleavage site: Arg-11e (autolytic) #status predicted
 F:483, 532, 633/Active site: His, Asp, Ser #status predicted

Query Match 10.1%; Score 475; DB 1; Length 686;
 Best Local Similarity 25.4%; Pred. No. 2, 3e-23;
 Matches 185; Conservative 88; Mismatches 275; Indels 180; Gaps 30;
 QY 227 RFTTPGPPDPYPAHARCQALRGDADSVSLTFRSPDLASCDERGDITVYNTLSPEM 286
 DB 29 RLASPGFP-GEVANDQERRWTLTAPGGRRLVFTHEDELSHLCEYDFVLSGAK-- 84
 QY 287 PHALVOLCG-----TYPSPYNLTFHSSQNVLLITLITRERHP--GFEATFQLPMS 339
 DB 85 --VLATLCGQESDTERAPKQDFFYLSGSLDITFRSDYSENEKPTGFEA-FYAAEDIDE 141
 QY 340 C-----GGLRAQGTFSNPPY 356
 DB 142 QVAPGEAPTCDHCHNLGFGYCSRAGVVLHRRKRTSCALSGVFTQSGELSSPEY 201
 QY 357 PGHYPPIIDCTWNIENVNNOHVAVKRFKFFYLRLPVPAGCPDQVYINEK-----YCG 411
 DB 202 PRPYPKLSSCTYSISLEGFSVILDFESPDVTH--PETLCPPVDFLKIQTDRHEHFGC 260
 QY 412 ERSQFVYTSNKNITVAFHSDQSYTDGFLAEVLSYSDSPCP-----GQFTCRTRGCR 466
 DB 261 KTLPHRLTETSNVTYITFVDESQDHTGKINHTSTAHACPFYMAPRNGVSVQAKYL 320
 QY 467 KEI.RCDGMADCTHSDDELNCSCDAGHQFTCKNFKCPFLFVVCDSVNDGDNDSQ--GC 523
 DB 321 K-----DSFISFCETGYELGCHPLKLSFTAVCK-----DGSMDRMPAC 361
 QY 524 SCRAQTRRCNGKCLSKQOCNGKDCGSDSDASCPKVVVYCTGHT-----YRCLN 576
 DB 362 SIY-----DCGP--PDLPSGRVYITGPGVYTKAVIYQSCER 398
 QY 577 GLCLSKNGPECDKEDC-SDG-----SDEKD--CD--CGLRSFTROARVVGSTDADEGR 625

A>Title: Genomic amplification with transcript sequencing.
 A:Reference number: 159529; MUID:88127096; PMID:3340835
 A:Accession: 159529
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 290-359 <RE2>
 A:Cross-references: GB:M19063; NID:g182622; PIDN:AAA52456.1; PID:g182623
 R:Agarwala, K.L.; Kawabata, S.; Takao, T.; Murata, H.; Shimonishi, Y.; Nishimura, H.; Iw
 Biochemistry 33, 5167-5171, 1994
 A>Title: Activation peptide of human factor IX has oligosaccharides O-glycosidically lin
 A:Reference number: A54255; MUID:94227047; PMID:8172892
 A:Accession: A54255
 A:Molecule type: protein
 A:Residues: 'D', 204, 'X', 206-211, 212, 'D', 214, 'X', 216-221, 'D', <AGA>
 A>Note: the residues designated 'X' were determined to be threonine bound to carbohydrate
 R:Di Sciopio, R.G.; Kurachi, K.; Davie, E.W.
 J. Clin. Invest. 61, 1528-1538, 1978
 A>Title: Activation of human factor IX (Christmas factor).
 A:Reference number: A18483; MUID:78194509; PMID:659613
 A:Contents: annotation; activation; active site; carbohydrate binding
 R:McGraw, R.A.; Davis, L.M.; Noyes, C.M.; Graham, J.B.; Roberts, H.R.; Stafford, D.W.
 Am. Soc. Hematol. Abstr. 64 (Suppl. 1), 262a, 1984
 A:Reference number: A37569
 A:Contents: annotation
 A>Note: 194-Thr was also found
 R:Morita, T.; Isaacs, B.S.; Esmen, C.T.; Johnson, A.E.
 J. Biol. Chem. 259, 5698-5704, 1984
 A>Title: Derivatives of blood coagulation factor IX contain a high affinity Ca2+-binding
 A:Reference number: A37543; MUID:84185715; PMID:6425296
 A:Contents: annotation; calcium binding
 R:Morita, T.; Isaacs, B.S.; Esmen, C.T.; Johnson, A.E.
 J. Biol. Chem. 260, 2583, 1985
 A:Reference number: A37544
 A:Contents: annotation; calcium binding, correction
 R:Bentley, A.K.; Reed, D.J.G.; Rizza, C.; Brownlee, G.G.
 Cell 45, 343-348, 1986
 A>Title: Defective propeptide processing of blood clotting factor IX caused by mutation
 A:Reference number: A37545; MUID:86189947; PMID:3009023
 A:Contents: annotation; signal sequence cleavage site
 R:Stehiro, K.; Kawabata, S.I.; Miyata, T.; Takeya, H.; Takamatsu, J.; Ogata, K.; Kamiya,
 J. Biol. Chem. 264, 21257-21265, 1989
 A>Title: Blood clotting factor IX B(M) Nagoya: substitution of arginine 180 by tryptophan
 A:Reference number: A30622; MUID:90078229; PMID:2592373
 A:Contents: annotation; sequence of mutant B(M) Nagoya
 A>Note: carboxylation, glycosylation, and cleavage sites
 R:Baron, M.; Norman, D.G.; Harvey, T.S.; Hanford, P.A.; Mayhew, M.; Tse, A.G.D.; Brownle
 submitted to the Brookhaven Protein Data Bank, November 1991
 A:Reference number: A51252; PDB:1IXA
 A:Contents: annotation; conformation by (1)H-NMR, residues 92-130
 A>Note: recombinant form expressed in yeast
 C:Comment: Factor IX is activated by factor Xla, which excises the activation peptide pr
 C:Comment: The gamma-carboxyglutamic acid residues arise by posttranslational, vitamin K
 C:Comment: Calcium binds to the gamma-carboxyglutamic acid (Gla) residues and, with strc
 C:Genetics:
 A:Gene: GDB:F9
 A:Cross-references: GDB:119900; OMIM:306900
 A:Map position: Xq27.1-Xq27.2
 A:Functions: 30/1; 84/2; 93/1; 131/1; 174/1; 241/3; 280/1
 C:Function:
 A:Description: catalyzes the proteolytic activation of coagulation factor X in the prese
 A:Pathway: blood coagulation intrinsic pathway
 C:Superfamily: coagulation factor X; EGF homology; Gla domain homology; trypsin homology
 C:Keywords: beta-hydroxyaspartic acid; blood coagulation; calcium binding; carboxyglutan
 F:1-28/Domain: signal sequence #status predicted <SIG>
 F:29-46/Domain: propeptide #status experimental <PP>
 F:31-91/Domain: Gla domain homology <GLA>
 F:47-119/Product: coagulation factor IXa light chain #status experimental <ALC>
 F:97-192/Domain: EGF homology <EG>
 F:134-170/Domain: EGF homology <EG2>
 F:192-226/Domain: activation peptide #status experimental <ACT>
 F:227-461/Product: coagulation factor IXa heavy chain #status experimental <AHC>
 F:227-454/Domain: trypsin homology <TRY>
 F:53,54,61,63,66,67,72,73,76,79,82,86/Modified site: gamma-carboxyglutamic acid (Glu) #

F:64-69,97-108,102-117,119-128,134-145,141-155,157-170,178-335,252-268,382-396,407-435/D/1
 F:99/Binding site: carbohydrate (Ser) (covalent) #status experimental
 F:110/Modified site: erythro-beta-hydroxyaspartic acid (Asp) #status experimental
 F:151-192/Cleavage site: Arg-Ala (coagulation factor Xla) #status experimental
 F:203,213/Binding site: carbohydrate (Asn) (covalent) #status experimental
 F:205,215/Binding site: carbohydrate (Thr) (covalent) #status experimental
 F:226-227/Cleavage site: Arg-Val (coagulation factor Xla) #status experimental

Query Match 10.1%; Score 474.5; DB 1; Length 461;
 Best Local Similarity 28.5%; Pred. No. 1.0e-23;
 Matches 139; Conservative 66; Mismatches 169; Indels 113; Gaps 19;

Qy	399	KDYVEINGEKYC-GERSQFVNTSNKITYRFRPSDQSYTPGFLAEIYSDSDPCPGQF	457
Db	53	EEFVQGLUEBCEBKCSFEAREVENTR-----TTERWKQVSDQCESNP----	101
Qy	458	TCRTGRGIRKELKCDGWADCTDHSDELNCSQAGHQ-----FTC--KXKFCPLFWVC	508
Db	102	-----CLNGSCKDIDINSYECPCPGFEKNCLELDTGKIKGRCEQF----	144
Qy	509	DSVNDGDNDEQ-GGSC-----PAQTFRCNGKCLSKSQCGKDKDCGDG	553
Db	145	-----CNNSADKNVVCSTEGYRLAENQSCBEAVPPCGRVSVQTSKLTTRA-----	192
Qy	554	SDASCPKVVVCTKRTYRCLNGLCLSKGNPCDCKEDSDGSDCKDCGHRSPTRQA	613
Db	193	--EAVFPDVYVAST-----AETLDNITQ-----SQSFNDFT	225
Qy	614	RVVGGTDADGEWPMQVSLHALQGHICGASLISPMWLSAHCYIDRGFRYSDDPTQT	673
Db	226	RVVGGEDAKGQGFPMQVVLNGKVD--FCGGSIVNEKIVTAACV--ETGVK1-----T	276
Qy	674	AFGLHDQSRASAPGVQERLKLIIHPFND--TFVDVIALLEKPAEYSMPVPLC	721
Db	277	VVAGEHN-IEETHTGKRVIRIIPHNNAIINKYNDHIALLEDEPLVLSVTPIC	335
Qy	732	LPDASHV-----PPAKAIWVTGMGHTQVGTGALLQKSEIRIVNOTTCENLLPQDITP	786
Db	336	IADKEYNITLTKGSG--YVSGMGRVFNHGRSALVQYRLVLDVRLATLRSRTKFTYN	392
Qy	787	RMCVGFSGVDSVDSGSPISVADGRITGAGVSWGDCGACGRNKGVTTLPLFR	846
Db	393	NMCAGFHGEGRSDSCQDSGSP-HVTEVESTFLTGIIISWEGCAMKGYITKVSRYV	451
Qy	847	DWIKENT 853	
Db	452	NWIKENT 458	

RESULT 26
 KFHHL
 coagulation factor Xla (EC 3.4.21.27) precursor [validated] - human
 N/Alternate names: antithrombolytic factor C; plasma chromoplasmin antecedent
 C/Species: Homo sapiens (hmn)
 C/Date: 13-Aug-1986 #sequence revision 26-May-1994 #text_change 09-Jul-2004
 C/Accession: A27431; A00920; A37940
 R:Aakari, R.; Davie, E.W.; Chung, D.W.
 Biochemistry 26, 7221-7228, 1987
 A>Title: Organization of the gene for human factor XI.
 A:Reference number: A27431; MUID:88107663; PMID:2827746
 A:Accession: A27431
 A:Molecule type: DNA
 A:Residues: 1-625 <AS>
 A:Cross-references: UNIPROT:P03951; GB:M18295
 A>Note: the sequence shown follows the authors' translation
 R:Fujikawa, K.; Chung, D.W.; Hendrickson, L.E.; Davie, E.W.
 Biochemistry 25, 2417-2424, 1986
 A>Title: Amino acid sequence of human factor XI, a blood coagulation factor with four tar
 A:Reference number: A00920; MUID:86243360; PMID:3636155
 A:Accession: A00920
 A:Molecule type: mRNA
 A:Residues: 1-625 <PU>
 A:Cross-references: GB:M13142; NID:g182832; PIDN:AAA52487.1; PID:g182833

R;McMullen, B.A.; Fujikawa, K.; Davie, E.W.
Biochemistry 30, 2056-2060, 1991
A;Title: Location of the disulfide bonds in human coagulation factor XI: the presence of
A;Reference number: A37940; MUID:9115017; PMID:1998667
A;Accession: A37940
A;Molecule type: protein
A;Residues: 28-33/35-49, 'X', 51-55, 'X', 57-63/70-75, 'X', 77-79/107-109, 'X', 111-112/132-139,
'X', 280-282, 'X', 284/285-297/313-316, 'X', 318-319/320-326, 'X', 328-330/347-349/373, 'X', 379
C;Comment: The proenzyme consists of two identical chains linked by one or more disulfide
he active site, and a heavy chain, which associates with high molecular weight (HMW) kin
C;Genetics:
A;Gene: GDB:F11
A;Cross-references: GDB:119891; OMIM:264900
A;Map position: 4q35-4q35
A;Intons: 19/1; 109/1; 162/2; 199/1; 252/2; 289/1; 343/2; 379/1; 435/2; 494/1; 52
C;Function:
A;Description: catalyzes the proteolytic activation of coagulation factor IX
A;Pathway: blood coagulation intrinsic pathway
C;Superfamily: coagulation factor XI; trypsin homology
C;Keywords: blood coagulation; duplication; glycoprotein; hemophilia C; homodimer; hydric
F;1-18/Domain: signal sequence #status predicted <SIG>
F;19-387/Product: coagulation factor Xla heavy chain #status experimental <HCH>
F;19-108/Domain: apple repeat <AP1>
F;109-198/Domain: apple repeat <AP2>
F;199-288/Domain: apple repeat <AP4>
F;290-379/Domain: apple repeat <AP5>
F;388-625/Product: coagulation factor Xla light chain #status experimental <LCH>
F;388-618/Domain: trypsin homology <TRY>
F;20-103/514-581, 571-599/Disulfide bonds: #status predicted
F;28/Disulfide bonds: interchain #status experimental
F;46-76, 50-56, 110-139, 136-165, 140-146, 200-283, 226-255, 230-236, 291-374, 317-346, 321-327, 38
F;90, 166, 353, 450/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;339/Disulfide bonds: interchain #status predicted
F;387-388/Cleavage site: Arg-116 (coagulation factor XIla) #status experimental
F;431, 480, 575/Active site: His, Asp, Ser #status predicted
F;491/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 10.1%; Score 471; DB 1; Length 625;
Best Local Similarity 35.4%; Pred. No. 3.8e-23;
Matches 118; Conservative 42; Mismatches 117; Indels 56; Gaps 13;

QY 554 SDEASCPKV--NVTCTKHYRLCLNGCLSKNCECGKEDC-----SDSDER----- 600
DB 313 SHBA-CQKLTNAKRCFFTYTPAQASC-----NGKKCKYLKLSNNSPTTLILGRG 364
QY 601 -----DCCGLRSFTROARVVGTTADDEGMPQVSHALG--QGHICGASLI 646
DB 365 GISGYTLRLCKMNECTTKI---KPRIVGTAASVSGMPQVTLHTSPQRHLGGSI 421
QY 647 SPNWLVAHACYIDDRGFRYSPTQWTAFLGLHDQO---RSAPGVQERRLRKRTISHP 702
DB 422 GNOMILTAHCF-----YGVESPRLILRVSGILNQSIKEDTSPFVQE-----ILTHDQ 471
QY 703 FNDFTPYDIALLELEKPAEYSSWVRICLPDASHVPRPAGKAIWYMGHTQYGGTAL 762
DB 472 YKMAESGYDIALKLETTVYVTSQRIICPSKGRNVIYTTDCWVYTWGGRKADKIQNT 531
QY 763 LQGEIRVINQTCSE--NLTPQOITPRMVCVFLSGVDSOCQSGGFLSGVEADGRIFG- 820
DB 532 LQAKIPIVYNBECQKRYRKHKITHKIKICAGYREBGDACKGDSGGFLSC--KENEVNHL 569
QY 821 AGVAVSWGCGAGNKKPGVYTRLPFLRDIKENT 853
DB 590 VGITSWGCGAQRERPGVYTNVVEYVDMLEKT 622

RESULT 27
S40007
trypsin (EC 3.4.21.4) precursor - African malaria mosquito
C;Species: Anopheles gambiae (African malaria mosquito)
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 09-Jul-2004
C;Accession: S40007
R;Muller, H.

submitted to the EMBL Data Library, June 1993
A;Reference number: S40003
A;Accession: S40007
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-275 <MOE>
A;Cross-references: UNIPROT:P35037; EMBL:Z22930; NID:G410323; PIDN:CAA80517.1; PID:G41032;
C;Superfamily: trypsin; trypsin homology
C;Keywords: hydrolase; protein digestion; serine proteinase
F;49-269/Domain: trypsin homology <TRY>
F;89, 134, 230/Active site: His, Asp, Ser #status predicted

Query Match 10.0%; Score 469; DB 2; Length 275;
Best Local Similarity 39.9%; Pred. No. 2e-23;
Matches 97; Conservative 39; Mismatches 89; Indels 18; Gaps 6;

QY 614 RVVCGTDADGEPMQVSHALGQGHICGASLISPMWLVSAHACYIDDRGFRSDPTQWT 673
DB 48 RIVGGFEIDVSETPYVSLQYF--NSHRCQGSVLSKMLITPAHCTVN-----LQPSILA 100
QY 674 AFLGLHDQSORSAPOVERLRKRIISHPFNDFTPYDIALLELEKPAEYSSWVRICLP 733
DB 101 VRLG-----SSRHASGCTVVRVARVLEHRYVDDSTIDYDSIMLEFTELTFSDVVGVSLE 156
QY 734 DASHVPRPAGKAIWYMGHTQYGGTALILQGEIRVINQTCENILPQO--ITPRMVCV 791
DB 157 EQDEAVEDGTMTVTVSQMGVQVSAESNAILIRANIPVNAQECTIAYSSSGGITDMLCA 216
QY 792 GFISGGVDSOCQSGGFLSGVEADGRIFGAGVSWGDCGAGNKKPGVYTRLPFLRDMKE 851
DB 217 GYRGGKDKACQDSGGL--VVDGKL--VGVSWGFGCAMPYPGVYAVAVNRVWRB 271
QY 852 NTG 854
DB 272 NSG 274

RESULT 28
146260
plasma (EC 3.4.21.7) precursor - western European hedgehog
C;Species: Erinaceus europaeus (western European hedgehog)
C;Date: 21-Feb-1997 #sequence_revision 21-Feb-1997 #text_change 09-Jul-2004
C;Accession: I46260
J;Lawn, R.M.; Boonmark, N.W.; Schwartz, K.; Lindahl, G.E.; Wade, D.P.; Byrne, C.D.; Fong,
J. Biol. Chem. 270, 24004-24009, 1995
A;Title: The recurring evolution of Lp(a): Insights from cloning of hedgehog apolipoprote
A;Reference number: I46259; MUID:96025778; PMID:7592597
A;Accession: I46260
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-810 <LAW>
A;Cross-references: UNIPROT:Q29485; EMBL:U33171; NID:G1046360; PID:G1046361
C;Superfamily: plasma; kringie homology; plasminogen-related protein precursor homology;
C;Keywords: hydrolase; serine proteinase
F;1-96/Domain: plasminogen-related protein precursor homology <PLPH>
F;103-181/Domain: kringie homology <KR1>
F;185-262/Domain: kringie homology <KR2>
F;275-352/Domain: kringie homology <KR3>
F;379-456/Domain: kringie homology <KR4>
F;482-561/Domain: kringie homology <KR5>
F;582-803/Domain: trypsin homology <TRY>

Query Match 10.0%; Score 468; DB 2; Length 810;
Best Local Similarity 28.1%; Pred. No. 8e-23;
Matches 158; Conservative 62; Mismatches 192; Indels 150; Gaps 27;

QY 405 NGEKTCGERSQGVYVTSNSKITYRFSHDS-----YDTGFLAFLYSDSDPCP 454
DB 279 NGEHYQGN-----VAVTVGLTCQRFWGEQSPRHDRTPENRYCKNLDENYCRNPDEPAP 333
QY 455 GQFTCTGTCIRKEL-----RCDSGMADCTPHSD-----LNGSCDAGHOF----- 494
DB 334 WCFT--TNSSVRWFCKIPDCVSSASETHSDAPVIVPEQTPVVOECTQAGNQTIRGTS 391

QY 495 --TCNKCKPLFV-----CD-SVNDGDNDEQSC---PAQTR 531
Db 392 STTYGKKCP--WTSRPHRSHSTPENYPADLTMYCNRPDGDGKMCYTTDPSRWE 449
QY 532 -CSNGKCL-----SKSQOC---NGKDDCGDSEAS----- 558
Db 450 FCNKKCSGTEMATNNSPVGVSSASBBDCLIDNGKGRGKATGAGTPOQAMAQ 509
QY 559 -----CPKVVVCTKATYRC-----LNGCLSKNPE-----CDGKEDCSGSD 599
Db 510 EPHRHSIFTPETINRADLOENY--CRNPDGANGPCYTTNPRKLFYCDIPIHCYSPSS-- 566
QY 600 KDCDCCGARSFTROARV-----VGGTDADGEMPMQVSLHALGOGHICGASLISPNWL 651
Db 567 -ADCG-----KKEVKKCCCPGRVGGCVAHPSMPQVSLRRPQ--HFCGGLIISPMV 617
QY 652 VSAHCYIDRGFFSDPTQWTAFLGLHDGQSRAPQVQERRLKRIISHPFNDFTPDYD 711
Db 618 VTAHCL-----EKFSPAIKVVLGAM--QETRLERVDQIKGVTKMFLP-----RAD 665
QY 712 IALIEKPAVSSMVRPCLPDASHVFPAGKAIWVGHTQYGGTGALILQGEIRVI 771
Db 666 IALLKLSPAIITDKHPACLPNSNMVWADRSCLYITGWEETK--GYGAGILKEAQLPVI 724
QY 772 NQTTCC--ENLLPQOITRRMVCVGLSGVDSGCGDSCGPLSSVEADRIFGAGVSMGD 829
Db 725 ENKYCNROSFLNGVNRSTELCAGHLAGVDSGQDSGPLVCFERD--RYIIQGVTSWGLG 783
QY 830 CAGNKKPGVYTRLLPFRDWIKR 851
Db 784 CARLTRGVYVRVRSYSLQD 805

RESULT 29

A57096
nudel protein precursor - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C>Date: 27-Oct-1995 #sequence_revision 27-Oct-1995 #text_change 09-Jul-2004
C:Accession: A57096
R:Hong, C.C.; Haehnleto, C.
Cell 82, 785-794, 1995
A:Title: An unusual mosaic protein with a protease domain, encoded by the nudel gene, is
A:Reference number: A57096, MUID:95401268; PMID:7671306
A:Accession: A57096
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-2616 <HON>
A:Cross-references: UNIPROT:P96159; GB:U29153; NID:9984320; PID:AAA83086.1; PID:9984321
C:Genetic8:
A:Gene: FlyBase:nd1
A:Cross-references: FlyBase:FBgn0002926
C:Superfamily: nudel protein; LDL receptor ligand-binding repeat homology; trypsin homol
C:Keywords: extracellular protein
F:891-929/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F:1145-1378/Domain: trypsin homology <TRY>
F:1196-1430/Domain: LDL receptor ligand-binding repeat homology <LDL7>
F:1776-1811/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F:2310-2344/Domain: LDL receptor ligand-binding repeat homology <LDL8>
F:2351-2387/Domain: LDL receptor ligand-binding repeat homology <LDL5>
F:2421-2457/Domain: LDL receptor ligand-binding repeat homology <LDL3>

Query Match 10.0%; Score 467.5; DB 2; Length 2616;
Best Local Similarity 22.9%; Pred. No. 3.3e-22;
Matches 210; Conservative 104; Mismatches 308; Indels 295; Gaps 41;

QY 97 TNEFNVAVENS-----NSTEFVSLASKVQDALKLTVSGVFLDPYKESAVT 144
Db 601 SSETVTTCYERQWLKKFEDARPQNELITPTGALDAKLDKKG-PKIPFLN----- 653
QY 145 AFSEGVYAYVWSFSLPQHLVEEAERVMAERVVMLEPPARSLKSFVVTSVVAFPTDSK 204
Db 654 -----GHTWN-----AADAQLSLCEBAL--RMR-----NKVATMSDGE 686

QY 205 TYOR-----TQDNCSEFGLHARGVELMRFTTPGPPDSPYPAAHARQMALRGDADSLTTF 260
Db 687 TKEGETFTTSPVQFTSRAPG-----GPPVSE-----TKKSAQCMFPPNF 729
QY 261 RSPDLASCDERGDLYVTYNTLSPMEPHALVOLGTYPPSYNTLPHSSQVNLITLITNT 320
Db 730 GMPSEIPVC-----FYMTPAIFRPMWMT-----PTMGQGAHFGSSN----- 769
QY 321 ERHHPGEALFPQLPN-----SSCGRLRKAGQTFNSPYTYGHYPRNIDCT 367
Db 770 ---PG--AaIFVPPQFGPSGNFPFGSGSGAGGQGANIFSKNASQKPTNQOQOYCS 823
QY 368 MNIEVPMNQGKVRKFKFYLLEPGVPAGTCKPVVEINGEKCYGERSQFVTSNSKITY 427
Db 824 Y---MNO-----SGRGAGS-----QTSQQQGGGSAF--SNAN-FKM 857
QY 428 RPHSDGSYTD-----TGFLAEYLSYDSDPCCQFTC--RTGRCIRKELRCDMA 475
Db 858 R-HANQSTANQOQIIVASYAGLPQPIQERSRCPEPDQFSFGQGECLPAARMCDNV 916
QY 476 DCTDHSBELNCSCDAGHFTCKKFKCPLFWVCDVNDCCGDNDEQSC--SCPAOTFPC-- 532
Db 917 DCSGSDSAC-----TCADRVBER--LCDGYEDCPMGEBELGCFGCESLAYSCE 966
QY 533 -----SNGKLSKSQCCNGKDCGSDGSEASCPKVVVYCTK----- 569
Db 967 NPQDFAKRNSTISMCTSRLEKCDGFNCLANGDEBCSN--LVTVDADHSHGASASEG 1024
QY 570 ---HTYR-----CLNG-----LCLSKNPECDGKEDCS----- 594
Db 1025 YIVHNYRGDWHPCNNNEKMAALACQMDENSRDHASLNVQTLTLPGFTEPSLHAG 1084
QY 595 -----DGSDEKD-----CDGCL-----RSFTQAR 614
Db 1085 VHFQAQCHGNNSHSLVDHYAVYKCPMOQGLPSKSMLESHSKVRRAVDSKRIVGDGR 1144
QY 615 VGGTDADGEMPMQVSLHALGOGHICGASLISPNMLVSAHCYIDRGFRYSDPTQWTA 674
Db 1145 IVGSHTSALQMPFVVAIYRNGKH--CGGITYISDRMTISAHCYINIGKTFYE-----V 1197
QY 675 FLGLHDGQSRAPQVQERRLKRIISHPFNDFTPDYDIALIEKPAVSSMVRPCLPD 734
Db 1198 RAGILRRSSYS--PATQIQPVSHVYVHQAYSRRSRNDLSLRLLNPLQFRMVRVPCLPD 1256
QY 735 ASHVFPAGKAIW-----VTGWHGTQYGGTGALILQSG-----EIRVINQTCEN 778
Db 1257 KGRITVGDMDWIMQVBEHTLCTVVGW-----AIREKGPSBPPMRQVIVPIRKCTD 1307
QY 779 LLPQOITRRMVCVGLSGVDSGCGDSCGPL--SYEADGRIFQAGVSMGDCAGRNKP 836
Db 1308 --PEDQASEDIQADDPGGRDACCQDGGGLFCRSVSNADDFYLAGVSHGNGCARQER 1365
QY 837 GVVYTRLLPFRDWIKENT 853
Db 1366 GVTYTRVTLVLDLEMAT 1382

RESULT 30

A30351
coagulation factor IXa (BC 3.4.21.22) precursor - dog
C:Species: Canis lupus familiaris (dog)
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C:Accession: A30351; I46201
R:Evans, J.P.; Walzke, H.H.; Ware, J.L.; Stafford, D.W.; High, K.A.
Blood 74, 207-212, 1989
A:Title: Molecular cloning of a cDNA encoding canine factor IX.
A:Reference number: A30351, MUID:89323338; PMID:2752110
A:Accession: A30351
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-452 <EVA>
A:Cross-references: UNIPROT:P19540; GB:M21757; NID:97972719; PID:AAA75006.1; PID:9163948

R; Axelrod, J.H.; Read, M.S.; Brinkhouse, K.M.; Verma, I.M.
Proc. Natl. Acad. Sci. U.S.A. 87, 5173-5177, 1990
A; Title: Phenotypic correction of factor IX deficiency in skin fibroblasts of hemophilic
A; Reference number: 146201; MUID: 90311364; PMID: 2367529
A; Accession: 146201
A; Status: preliminary; translated from GB/EMBL/DBJ
A; Molecule type: mRNA
A; Residues: 1-452 <AXE>
A; Cross-references: GB:M33826; NID:g163949; PIDN:AAA30844.1; PID:g163950
C; Superfamily: coagulation factor X; EGF homology; Gla domain homology; trypsin homology
C; Keywords: beta-hydroxyaspartic acid; blood coagulation; calcium binding; carboxylatan
F; 1-21/Domain: signal sequence #status predicted <SIG>
F; 22-40/Domain: propeptide #status predicted <PRO>
F; 41-452/Product: coagulation factor IX #status predicted <MAT>
F; 90-121/Domain: EGF homology <EG1>
F; 127-163/Domain: EGF homology <EG2>
F; 218-445/Domain: trypsin homology <TRY>
F; 46, 47, 54, 56, 59, 60, 65, 66, 69, 72, 75, 79/Modified site: gamma-carboxylglutamic acid (Glu) #
F; 57-62, 90-101, 95-110, 112-121, 127-138, 134-148, 150-163, 171-326, 243-259, 373-387, 398-426/D1
F; 258, 306, 402/Active site: His, Asp, Ser #status predicted

Query Match 10.0%; Score 466.5; DB 1; Length 452;
Best Local Similarity 28.5%; Pred. No. 5.1e-23;
Matches 140; Conservative 63; Mismatches 164; Indels 125; Gaps 19;

QY 399 KDVEINGEKYC-GERSQF---VVTNSNKITVRFHSDQSYDTGFLAEYLSDSSDPC 453
Db 46 EEFVRGLERECIEKCSFEAREVFENTK-----TTFWKQYVDGQ----- 89

QY 454 PGQFTCTGRCIRKELRCGMDCTDHSDELNCSDAHQFTCKNFKCPPLFWVDSVND 513
Db 90 -----CESNPCLN-----DG--VCKDDINSYECWCRAFG----- 116

QY 514 CGNSDDEQSGSCPAQ-TFRCSNGKCLSKQCCNGKDDGSDGSDASCPKNVVTCTKTY 572
Db 117 -----EGKNCBLDVTGNIRKNC-----KQFCKLGP-----NKVCS--- 149

QY 573 RCLNGCLSKGNPECD-----GKEDCSGSD-EKDCDCG 605
Db 150 -CTTGYGLADQSCBPAVFPQGRVSVPHISMRTBAETLFSNMDYENSTVEKILDNV 208

QY 606 LRSEFTQARVVGTDADGEMPMQVSLHALGQGHICGASLISPMVLVSAHCYIDRGR 665
Db 209 TQPLNDFTRVVGKDAKPGQFPWQVLLNGKVA-FCGGSIIINEKVVTAACIEPD---- 263

QY 666 YSDPTQWTAFLGLHDSQSRAPGQERLRKLIISHPENF--TFQDIALLEEKPAEY 723
Db 264 ----VKTIIVAGEHNTKREHT-BQKKNVIRTIILHSHSYNATINKYNDIALLEIDEP 318

QY 724 SSMVRPILCLDA-SHVFPAKAIWVTGWHGTQYGTGALILQGEIRVINQTCENILP 781
Db 319 NSYVTPICIADRESNIFLKFSGSYVSGWGRVFNKGSASILQYLKPLVDRAATCLASTK 378

QY 782 QQITPRMCTVGLSGVDSQGDSSGGLSSVEADGRIFGAGVWSMGDCAGRNKPGYTR 841
Db 379 FTIYNNMFCAGFHGKDSQGDSSGP-HYTEVEGISFLTGIIISWBECAKMGKYGITYK 437

QY 842 LPLPRDMIKENT 853
Db 438 VSRVNMWIKET 449

Search completed: November 29, 2004, 08:33:30
Job time : 39.6873 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 29, 2004, 08:32:39 ; Search time 111.183 Seconds
(without alignments)
2727.603 Million cell updates/sec

Title: US-09-936-333-27

Perfect score: 4683
Sequence: 1 MGSDBARKGGGCPKDFAGL.....PGVYTRLPFRDWIKENTGV 855

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1575965 seqs, 354694765 residues

Total number of hits satisfying chosen parameters: 1575965

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 100 summaries

Database :

Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubppaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubppaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubppaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubppaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubppaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubppaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubppaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubppaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubppaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubppaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubppaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubppaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubppaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubppaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubppaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pep.*
- 18: /cgn2_6/ptodata/2/pubppaa/US11_NEW_PUB.pep.*
- 19: /cgn2_6/ptodata/2/pubppaa/US60_NEW_PUB.pep.*
- 20: /cgn2_6/ptodata/2/pubppaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4659	99.5	855	US-10-295-027-1185	Sequence 1185, App
2	4659	99.5	855	US-10-072-012-353	Sequence 353, App
3	4659	99.5	855	US-10-072-012-412	Sequence 412, App
4	4659	99.5	855	US-10-072-012-419	Sequence 419, App
5	4655	99.4	855	US-10-072-012-354	Sequence 354, App
6	4655	99.4	855	US-10-072-012-420	Sequence 420, App
7	4655	99.4	855	US-10-037-012-132	Sequence 132, App
8	4653	99.4	855	US-09-776-191-132	Sequence 132, App
9	4653	99.4	855	US-10-099-700A-2	Sequence 2, App11
10	4653	99.4	855	US-10-190-030B-2	Sequence 2, App11
11	4653	99.4	855	US-10-302-840A-2	Sequence 2, App11
12	4653	99.4	855	US-10-267-219-2	Sequence 2, App11
13	4653	99.4	855	US-10-112-221A-2	Sequence 2, App11

14	4653	99.4	855	US-10-104-271-2	Sequence 2, App11
15	4653	99.4	855	US-10-147-211A-2	Sequence 2, App11
16	4653	99.4	855	US-10-156-214A-2	Sequence 2, App11
17	4653	99.4	855	US-10-072-012-352	Sequence 352, App
18	4653	99.4	855	US-10-072-012-411	Sequence 411, App
19	4653	99.4	855	US-10-072-012-418	Sequence 418, App
20	4653	99.4	855	US-10-600-187-2	Sequence 2, App11
21	4614	98.5	851	US-10-276-774-1798	Sequence 1798, App
22	4614	98.5	851	US-10-296-115-1143	Sequence 1143, App
23	4167	89.0	762	US-10-729-807-1	Sequence 1, App11
24	4158.5	88.8	782	US-10-097-340-312	Sequence 312, App
25	4092	87.4	757	US-10-072-012-44	Sequence 44, App11
26	3884	82.9	855	US-09-900-751-2	Sequence 2, App11
27	3884	82.9	855	US-10-072-012-355	Sequence 355, App
28	3884	82.9	855	US-10-072-012-413	Sequence 413, App
29	3866	82.6	855	US-10-072-012-356	Sequence 356, App
30	3866	82.6	855	US-10-072-012-414	Sequence 414, App
31	3866	82.6	855	US-10-072-012-417	Sequence 417, App
32	3793	81.0	902	US-10-333-743-3	Sequence 3, App11
33	3793	81.0	902	US-10-600-187-10	Sequence 10, App11
34	3793	81.0	902	US-10-297-987B-11	Sequence 11, App11
35	2986	63.8	620	US-09-925-301-1193	Sequence 1193, App
36	2660	56.8	845	US-10-072-012-415	Sequence 415, App
37	1305	27.9	241	US-09-776-191-50	Sequence 50, App11
38	1305	27.9	241	US-10-099-700A-4	Sequence 4, App11
39	1305	27.9	241	US-10-092-004A-2	Sequence 2, App11
40	1305	27.9	241	US-10-190-030B-4	Sequence 4, App11
41	1305	27.9	241	US-10-302-840A-4	Sequence 4, App11
42	1305	27.9	241	US-10-267-219-4	Sequence 4, App11
43	1305	27.9	241	US-10-112-221A-4	Sequence 4, App11
44	1305	27.9	241	US-10-104-271-4	Sequence 4, App11
45	1305	27.9	241	US-10-147-211A-4	Sequence 4, App11
46	1114.5	23.8	799	US-10-072-012-410	Sequence 410, App
47	1114.5	23.8	799	US-10-072-012-416	Sequence 416, App
48	1112	23.7	802	US-09-888-615-113	Sequence 113, App
49	1112	23.7	802	US-09-978-295A-169	Sequence 169, App
50	1112	23.7	802	US-09-978-697-169	Sequence 169, App
51	1112	23.7	802	US-09-978-192A-169	Sequence 169, App
52	1112	23.7	802	US-09-999-832A-169	Sequence 169, App
53	1112	23.7	802	US-09-978-189-169	Sequence 169, App
54	1112	23.7	802	US-09-978-608A-169	Sequence 169, App
55	1112	23.7	802	US-09-978-585A-169	Sequence 169, App
56	1112	23.7	802	US-09-978-191A-169	Sequence 169, App
57	1112	23.7	802	US-09-978-403A-169	Sequence 169, App
58	1112	23.7	802	US-09-978-564A-169	Sequence 169, App
59	1112	23.7	802	US-09-999-832A-169	Sequence 169, App
60	1112	23.7	802	US-09-981-915A-169	Sequence 169, App
61	1112	23.7	802	US-09-978-824-169	Sequence 169, App
62	1112	23.7	802	US-09-918-585A-169	Sequence 169, App
63	1112	23.7	802	US-09-999-834A-169	Sequence 169, App
64	1112	23.7	802	US-09-978-423A-169	Sequence 169, App
65	1112	23.7	802	US-09-978-193A-169	Sequence 169, App
66	1112	23.7	802	US-09-999-830A-169	Sequence 169, App
67	1112	23.7	802	US-09-978-757A-169	Sequence 169, App
68	1112	23.7	802	US-09-978-187B-169	Sequence 169, App
69	1112	23.7	802	US-09-978-643A-169	Sequence 169, App
70	1112	23.7	802	US-09-978-375A-169	Sequence 169, App
71	1112	23.7	802	US-09-978-298A-169	Sequence 169, App
72	1112	23.7	802	US-09-978-188A-169	Sequence 169, App
73	1112	23.7	802	US-09-978-681A-169	Sequence 169, App
74	1112	23.7	802	US-09-978-194A-169	Sequence 169, App
75	1112	23.7	802	US-09-999-829A-169	Sequence 169, App
76	1112	23.7	802	US-09-978-299A-169	Sequence 169, App
77	1112	23.7	802	US-09-978-544A-169	Sequence 169, App
78	1112	23.7	802	US-09-978-665A-169	Sequence 169, App
79	1112	23.7	802	US-09-978-802A-169	Sequence 169, App
80	1112	23.7	802	US-09-999-831A-169	Sequence 169, App
81	1112	23.7	802	US-10-017-081A-169	Sequence 169, App
82	1112	23.7	802	US-10-167-749-169	Sequence 169, App
83	1112	23.7	802	US-10-013-921A-169	Sequence 169, App
84	1112	23.7	802	US-10-013-929A-169	Sequence 169, App
85	1112	23.7	802	US-10-016-177A-169	Sequence 169, App
86	1112	23.7	802	US-10-166-709A-169	Sequence 169, App

```
87 1112 23.7 802 14 US-10-143-031A-169 Sequence 169, App
88 1112 23.7 802 14 US-10-143-030A-169 Sequence 169, App
89 1112 23.7 802 14 US-10-002-967A-169 Sequence 169, App
90 1112 23.7 802 14 US-10-017-083A-169 Sequence 169, App
91 1112 23.7 802 14 US-10-145-128A-169 Sequence 169, App
92 1112 23.7 802 14 US-10-017-191A-169 Sequence 169, App
93 1112 23.7 802 14 US-10-143-028A-169 Sequence 169, App
94 1112 23.7 802 14 US-10-143-029A-169 Sequence 169, App
95 1112 23.7 802 14 US-10-145-089A-169 Sequence 169, App
96 1112 23.7 802 14 US-10-165-067A-169 Sequence 169, App
97 1112 23.7 802 14 US-10-145-017A-169 Sequence 169, App
98 1112 23.7 802 14 US-10-164-728A-169 Sequence 169, App
99 1112 23.7 802 14 US-10-013-926A-169 Sequence 169, App
100 1112 23.7 802 14 US-10-165-247A-169 Sequence 169, App
```

ALIGNMENTS

```
RESULT 1
US-10-295-027-1185
; Sequence 1185, Application US/10295027
; Publication No. US2003023350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Nacasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynn, Richard
; APPLICANT: Hevez, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OR INVENTION: Methods of Diagnosis of Cancer, Compositions and
; FILE REFERENCE: 018501-012500US
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US/10/295, 027
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/363, 733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350, 666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335, 394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332, 464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334, 393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340, 376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347, 211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347, 349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355, 250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356, 714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1185
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-295-027-1185
```

```
Query Match 99.5%; Score 4659; DB 14; Length 855;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 852; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

1 MGSDDRARKGGGPKDPCAGLTKYNSRHEKVNGLBEGVEFLPVNNVKKVEKKGPRWVLLAA 60

```
|||||
1 MGSDDRARKGGGPKDPCAGLTKYNSRHEKVNGLBEGVEFLPVNNVKKVEKKGPRWVLLAA 60
61 VLLGLLVLLIGLGLVWHLQYRDPYVOKVNGVMRTTNENFVAYNSNSTEVSLSKV 120
61 VLLGLLVLLIGLGLVWHLQYRDPYVOKVNGVMRTTNENFVAYNSNSTEVSLSKV 120
121 KDLAKLVSQVPLGPGYHKESAVTAFSEGSVIAYYSEFS1PQHLVEAEERVAERVVM 180
121 KDLAKLVSQVPLGPGYHKESAVTAFSEGSVIAYYSEFS1PQHLVEAEERVAERVVM 180
181 LPFRASLKSFFVTSVVAFPPTDSKTYQRTQDNCSCGILHARGVLMRFTTPGPDSPYPA 240
181 LPFRASLKSFFVTSVVAFPPTDSKTYQRTQDNCSCGILHARGVLMRFTTPGPDSPYPA 240
241 HARCOMALRGDADSVLSLFRSFDLASCDERGSIDLTVYNTVLSPMEPHALVOLCGYTPPS 300
241 HARCOMALRGDADSVLSLFRSFDLASCDERGSIDLTVYNTVLSPMEPHALVOLCGYTPPS 300
301 YNLTFHSSQNVLLITLITNTERRHGFEATFFQLPRMSSCGRLRAQGTFFSPYGHY 360
301 YNLTFHSSQNVLLITLITNTERRHGFEATFFQLPRMSSCGRLRAQGTFFSPYGHY 360
361 PENIDCTWNI EVDNNOHVKRFKFFYLLEPGVPAGTCPKDYVEINBKTGGESEQVYVS 420
361 PENIDCTWNI EVDNNOHVKRFKFFYLLEPGVPAGTCPKDYVEINBKTGGESEQVYVS 420
421 NSNKITVRFPSSDQSYDTGTGLAEYLSYSSDPCPGQFTCRGTGCIKRELCDMACTPH 480
421 NSNKITVRFPSSDQSYDTGTGLAEYLSYSSDPCPGQFTCRGTGCIKRELCDMACTPH 480
481 SDELINCSDAQHOFCKNFKCKPLFWVCDSVNDGDNSEOGSCPAQTFRCNGKCLSK 540
481 SDELINCSDAQHOFCKNFKCKPLFWVCDSVNDGDNSEOGSCPAQTFRCNGKCLSK 540
541 SQCCNGKDCGDSDBASCPKXNVVTCTKTYRCLNGLCLSKNPECDGKEDSDSDSK 600
541 SQCCNGKDCGDSDBASCPKXNVVTCTKTYRCLNGLCLSKNPECDGKEDSDSDSK 600
601 DCCGGLRSFTROARVVGTDADGEMPMOVSLLHALGQHCASLISPMVLVSAACIYD 660
601 DCCGGLRSFTROARVVGTDADGEMPMOVSLLHALGQHCASLISPMVLVSAACIYD 660
661 DRGFRYSDFPTWTAFLGLHDQSGRSAPGVQERLKR1ISHPFNDFTFYDIALLEKRP 720
661 DRGFRYSDFPTWTAFLGLHDQSGRSAPGVQERLKR1ISHPFNDFTFYDIALLEKRP 720
721 AEYSSWVRPCLPDASHVFPAGKAIWVTGMGHTQYGGTGLALIQKBEIRVINTGENLL 780
721 AEYSSWVRPCLPDASHVFPAGKAIWVTGMGHTQYGGTGLALIQKBEIRVINTGENLL 780
781 PQGTRPMKCVGLSGGVDSGQSDSGPLSSVEADGRIFQAGVVSMDGCAQNNKGGVYT 840
781 PQGTRPMKCVGLSGGVDSGQSDSGPLSSVEADGRIFQAGVVSMDGCAQNNKGGVYT 840
841 RLPLFRDWIKENTGV 855
841 RLPLFRDWIKENTGV 855
```

```
RESULT 2
US-10-072-012-353
; Sequence 353, Application US/10072012
; Publication No. US20040033493A1
; GENERAL INFORMATION:
; APPLICANT: Tchernev, Velizar
; APPLICANT: Spytek, Kimberly
; APPLICANT: Zethusen, Bryan
; APPLICANT: Patuturajan, Meera
; APPLICANT: Shimkete, Richard
; APPLICANT: Li, Li
; APPLICANT: Gangolli, Bsha
; APPLICANT: Padigarau, Muralidhara
```

APPLICANT: Anderson, David W.
APPLICANT: Rastelli, Luca
APPLICANT: Miller, Charles E.
APPLICANT: Gerlach, Valerie
APPLICANT: Taupier Jr, Raymond J.
APPLICANT: Gusev, Vladimir Y.
APPLICANT: Coleman, Steven D.
APPLICANT: Wolenc, Adam R.
APPLICANT: Pena, Carol E. A
APPLICANT: Furtak, Katarzyna
APPLICANT: Grose, William M.
APPLICANT: Alsobrook II, John P.
APPLICANT: Lepley, Denise M.
APPLICANT: Rieger, Daniel K.
APPLICANT: Burgess, Catherine E.
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
FILE REFERENCE: 21402-258
CURRENT APPLICATION NUMBER: US/10/072,012
CURRENT FILING DATE: 2002-01-31
PRIOR APPLICATION NUMBER: 60/265,102
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: 60/265,514
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265,517
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265,412
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265,395
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/266,406
PRIOR FILING DATE: 2001-02-02
PRIOR APPLICATION NUMBER: 60/266,767
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: 60/267,057
PRIOR FILING DATE: 2001-02-07
PRIOR APPLICATION NUMBER: 60/266,975
PRIOR FILING DATE: 2001-02-07
PRIOR APPLICATION NUMBER: 60/267,459
PRIOR FILING DATE: 2001-02-08
Remaining Prior Application data removed - See file wrapper or PALM.
NUMBER OF SEQ ID NOS: 1391
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 353
LENGTH: 855
TYPE: PRT
ORGANISM: Homo sapiens
US-10-072-012-353
Query Match 99.5%; Score 4659; DB 15; Length 855;
Best Local Similarity 99.6%; Pred. No. 0; Matches 852; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

301 YNLFHSSQVLLITLITLINTERBRHGFEATFFQLPRMSSCGRLRKAQGTFSPPYPGHY 360
361 PPNIIDCTWNIIEVPPNNOVKVRFKFFYLLEPGVPAGTCKPYVLENGKTCGSEQFPVTS 420
361 PPNIIDCTWNIIEVPPNNOVKVRFKFFYLLEPGVPAGTCKPYVLENGKTCGSEQFPVTS 420
421 NSNKTIVRPHSDOSYTPGTGFLAEYLSYSDSDPCGQGTCTGRCIRKELACDGMADCTDH 480
421 NSNKTIVRPHSDOSYTPGTGFLAEYLSYSDSDPCGQGTCTGRCIRKELACDGMADCTDH 480
481 SDELNCSGDAHQPTCKNKECKPLFWVCDSVNDGSDSDGSCSPAQTRCNSGKCLSK 540
481 SDELNCSGDAHQPTCKNKECKPLFWVCDSVNDGSDSDGSCSPAQTRCNSGKCLSK 540
541 SQQCNKGDDGSDSDGSCSPAQTRCNSGKCLSK 600
541 SQQCNKGDDGSDSDGSCSPAQTRCNSGKCLSK 600
601 DDCGRLSFTROARVVGCTDADGEMPMOYSLHALGQGHICGASLISPMVLVSAHGYID 660
601 DDCGRLSFTROARVVGCTDADGEMPMOYSLHALGQGHICGASLISPMVLVSAHGYID 660
661 DRGFRYSPTQWTAFLGLHDSQSRAPGVQERRLKR11SHPPFNDFTFDYDIALLEKXP 720
661 DRGFRYSPTQWTAFLGLHDSQSRAPGVQERRLKR11SHPPFNDFTFDYDIALLEKXP 720
721 AEVSMWRP1CLPASHVFPAGKAIWTTGNGHIOYGGTGL11QKGBIRINOTTCNLL 780
721 AEVSMWRP1CLPASHVFPAGKAIWTTGNGHIOYGGTGL11QKGBIRINOTTCNLL 780
781 PQOITPRMCMCVFISGSDVSCQSDGSGPLSSVEADGRIIPAGVYVWGDGAGRRKPGVYT 840
781 PQOITPRMCMCVFISGSDVSCQSDGSGPLSSVEADGRIIPAGVYVWGDGAGRRKPGVYT 840
841 RLPLFRDMIKENTGV 855
841 RLPLFRDMIKENTGV 855

RESULT 3
US-10-072-012-412
Sequence 412, Application US/10072012
Publication No. US20040033493A1
GENERAL INFORMATION:
APPLICANT: Tchervetev, Velizar
APPLICANT: Spytek, Kimberly
APPLICANT: Zernhusen, Bryan
APPLICANT: Patuturajan, Meera
APPLICANT: Shinkels, Richard
APPLICANT: Li, Li
APPLICANT: Gangoli, Baha
APPLICANT: Padigaru, Muralidhara
APPLICANT: Anderson, David W.
APPLICANT: Rastelli, Luca
APPLICANT: Miller, Charles E.
APPLICANT: Gerlach, Valerie
APPLICANT: Taupier Jr, Raymond J.
APPLICANT: Gusev, Vladimir Y.
APPLICANT: Coleman, Steven D.
APPLICANT: Wolenc, Adam R.
APPLICANT: Pena, Carol E. A
APPLICANT: Furtak, Katarzyna
APPLICANT: Grose, William M.
APPLICANT: Alsobrook II, John P.
APPLICANT: Lepley, Denise M.
APPLICANT: Rieger, Daniel K.
APPLICANT: Burgess, Catherine E.
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
FILE REFERENCE: 21402-258
CURRENT APPLICATION NUMBER: US/10/072,012
CURRENT FILING DATE: 2002-01-31
PRIOR APPLICATION NUMBER: 60/265,102

;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: 60/265,514
;; PRIOR FILING DATE: 2001-01-31
;; PRIOR APPLICATION NUMBER: 60/265,517
;; PRIOR FILING DATE: 2001-01-31
;; PRIOR APPLICATION NUMBER: 60/265,412
;; PRIOR FILING DATE: 2001-01-31
;; PRIOR APPLICATION NUMBER: 60/265,395
;; PRIOR FILING DATE: 2001-01-31
;; PRIOR APPLICATION NUMBER: 60/266,406
;; PRIOR FILING DATE: 2001-02-02
;; PRIOR APPLICATION NUMBER: 60/266,767
;; PRIOR FILING DATE: 2001-02-05
;; PRIOR APPLICATION NUMBER: 60/267,057
;; PRIOR FILING DATE: 2001-02-07
;; PRIOR APPLICATION NUMBER: 60/266,975
;; PRIOR FILING DATE: 2001-02-07
;; PRIOR APPLICATION NUMBER: 60/267,459
;; PRIOR FILING DATE: 2001-02-08
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 1391
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 412
;; LENGTH: 855
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-072-012-412

Query Match 99.5%; Score 4659; DB 15; Length 855;
Best Local Similarity 99.6%; Pctid. No. 0;
Matches 852; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MGSDRAKGGGKPGDFAGLKYNSRHKVNGLEBGEVFLPVNNVKKYKKGPRGMVLLAA 60
DB 1 MGSDRAKGGGKPGDFAGLKYNSRHKVNGLEBGEVFLPVNNVKKYKKGPRGMVLLAA 60
QY 61 VLIGLLVLIGLIGFLVHMLQYRDVAVKXKGYRITNENFVDAYENSNSTEFLVSLASKV 120
DB 61 VLIGLLVLIGLIGFLVHMLQYRDVAVKXKGYRITNENFVDAYENSNSTEFLVSLASKV 120
QY 121 KDLAKLISYGVPELGPYHKSAAVAFSEGSVIATYMESEFIPQHLVEAEKRVMAEERVVM 180
DB 121 KDLAKLISYGVPELGPYHKSAAVAFSEGSVIATYMESEFIPQHLVEAEKRVMAEERVVM 180
QY 181 LPPRARSLSKSFVVVSVVAFPTDSKTVDRTDNSSCFGLHARGVLMKFTTGGFFDSYPYA 240
DB 181 LPPRARSLSKSFVVVSVVAFPTDSKTVDRTDNSSCFGLHARGVLMKFTTGGFFDSYPYA 240
QY 241 HARCQMLRGDADSVLSLTFRSPDLASCDEKSDLVTVYNTLSMEPHALVQLCGTYPSP 300
DB 241 HARCQMLRGDADSVLSLTFRSPDLASCDEKSDLVTVYNTLSMEPHALVQLCGTYPSP 300
QY 301 YNLFHSSQNVLLITLTNTERRHRRGFEATPFQLRPMSSCGGRIRKQGTFFNSPYRGHY 360
DB 301 YNLFHSSQNVLLITLTNTERRHRRGFEATPFQLRPMSSCGGRIRKQGTFFNSPYRGHY 360
QY 361 PPNIDCTWNTLEVPNNQHVAFKFFVLLBPGVAPGCTPKDQVVEINSGYSGERSQFVYTS 420
DB 361 PPNIDCTWNTLEVPNNQHVAFKFFVLLBPGVAPGCTPKDQVVEINSGYSGERSQFVYTS 420
QY 421 NSNKITVTRFHSDDSYTDTGFLAEYLSYDSDPCGQGTCTGRCIRKELRCDGADCTDH 480
DB 421 NSNKITVTRFHSDDSYTDTGFLAEYLSYDSDPCGQGTCTGRCIRKELRCDGADCTDH 480
QY 481 SDELNGSCDGHQGTCKNKKCPFLFWCDVNDGDGSDSCCPAOTFRCSNGKLSK 540
DB 481 SDELNGSCDGHQGTCKNKKCPFLFWCDVNDGDGSDSCCPAOTFRCSNGKLSK 540
QY 541 SQQNGKDDGSDGDEASCPKVVNVVCTKHTYRCLNGLCLSKNGPECDEKEDCSGSDSK 600
DB 541 SQQNGKDDGSDGDEASCPKVVNVVCTKHTYRCLNGLCLSKNGPECDEKEDCSGSDSK 600
QY 601 DCCGRLSFTRQARVVGCTDADGEMPMQVSLHALGGHICGASLISPMVLVSAHACYID 660

DB 601 DCCGRLSFTRQARVVGCTDADGEMPMQVSLHALGGHICGASLISPMVLVSAHACYID 660
QY 661 DRGFRYSDDPTQWTAFLGIDHOSORSAKGVORLKRIIISHPFNDFTFDIDLLELEKP 720
DB 661 DRGFRYSDDPTQWTAFLGIDHOSORSAKGVORLKRIIISHPFNDFTFDIDLLELEKP 720
QY 721 AEYSSWVRPCLDPAHVFPAGKAIWVTGNGHTQYGGTGLILQKGEIRVINTCCENLL 780
DB 721 AEYSSWVRPCLDPAHVFPAGKAIWVTGNGHTQYGGTGLILQKGEIRVINTCCENLL 780
QY 781 PQOITPRMVCVGLSGVDSCQSDSGPLSSVADRIIRAGVYVSGDGAGNKGYYT 840
DB 781 PQOITPRMVCVGLSGVDSCQSDSGPLSSVADRIIRAGVYVSGDGAGNKGYYT 840
QY 841 RLPLFRDWIKENTGV 855
DB 841 RLPLFRDWIKENTGV 855

RESULT 4
US-10-072-012-419
;; Sequence 419, Application US/10072012
;; Publication No. US20040033493A1
;; GENERAL INFORMATION:
;; APPLICANT: Tcheurev, Velizar
;; APPLICANT: Spytek, Kimberly
;; APPLICANT: Zethusen, Bryan
;; APPLICANT: Patturajan, Meera
;; APPLICANT: Shinkete, Richard
;; APPLICANT: Li, Li
;; APPLICANT: Gangoli, Esha
;; APPLICANT: Padigaru, Muraidhara
;; APPLICANT: Anderson, David W.
;; APPLICANT: Rastelli, Luca
;; APPLICANT: Miller, Charles E.
;; APPLICANT: Gerlach, Valerie
;; APPLICANT: Taupier Jr, Raymond J.
;; APPLICANT: Gusev, Vladimir Y.
;; APPLICANT: Coleman, Steven D.
;; APPLICANT: Wolenc, Adam R.
;; APPLICANT: Pena, Carol E. A
;; APPLICANT: Futrak, Katarzyna
;; APPLICANT: Grose, William M.
;; APPLICANT: Alsobrook II, John P.
;; APPLICANT: Lepley, Denise M.
;; APPLICANT: Rieger, Daniel K.
;; APPLICANT: Burgess, Catherine E.
;; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
;; FILE REFERENCE: 21402-258
;; CURRENT APPLICATION NUMBER: US/10/072,012
;; CURRENT FILING DATE: 2002-01-31
;; PRIOR APPLICATION NUMBER: 60/265,102
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: 60/265,514
;; PRIOR FILING DATE: 2001-01-31
;; PRIOR APPLICATION NUMBER: 60/265,517
;; PRIOR FILING DATE: 2001-01-31
;; PRIOR APPLICATION NUMBER: 60/265,412
;; PRIOR FILING DATE: 2001-01-31
;; PRIOR APPLICATION NUMBER: 60/265,395
;; PRIOR FILING DATE: 2001-01-31
;; PRIOR APPLICATION NUMBER: 60/266,406
;; PRIOR FILING DATE: 2001-02-02
;; PRIOR APPLICATION NUMBER: 60/266,767
;; PRIOR FILING DATE: 2001-02-05
;; PRIOR APPLICATION NUMBER: 60/267,057
;; PRIOR FILING DATE: 2001-02-07
;; PRIOR APPLICATION NUMBER: 60/266,975
;; PRIOR FILING DATE: 2001-02-07
;; PRIOR APPLICATION NUMBER: 60/267,459
;; Remaining Prior Application data removed - See File Wrapper or PALM.

```

; NUMBER OF SEQ ID NOS: 1391
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 419
; LENGTH: 855
; TYPE: PR1
; ORGANISM: Homo sapiens
US-10-072-012-419

```

Query Match	99.5%	Score 4659;	DB 15;	Length 855;
Best Local Similarity	99.6%	Pred. No. 0;		
Matches 852;	Conservative	0;	Mismatches 3;	Indels 0;
				Gaps 0;

Qy	MSGDRARKGGCGKJGDGAGLKTNSRHEKNGJLEBEVEFLPVNNYKXKJGKJGPGMWVTLAA	60
Db	1 MSGDRARKGGCGKJGDGAGLKTNSRHEKNGJLEBEVEFLPVNNYKXKJGKJGPGMWVTLAA	60
Qy	VJLGLLVJLJGJLGVMLQYEDVVRQVKXGMYRITMENFVDAYENSNTSEFVSLASKV	120
Db	61 VJLGLLVJLJGJLGVMLQYEDVVRQVKXGMYRITMENFVDAYENSNTSEFVSLASKV	120
Qy	KDMLKLLYSGVPLPGYHKESAVTPAFSEBSVTAAYWSEFS1POHLVEAEKVAEVEVVM	180
Db	121 KDMLKLLYSGVPLPGYHKESAVTPAFSEBSVTAAYWSEFS1POHLVEAEKVAEVEVVM	180
Qy	LPEPRASLSFVVTSVVAEPFDSKTVQRTQDSCSFGJHAAGVEMRPTTGPFPDSYPYPA	240
Db	181 LPEPRASLSFVVTSVVAEPFDSKTVQRTQDSCSFGJHAAGVEMRPTTGPFPDSYPYPA	240
Qy	HACQWALRGDADSVLSTPNSPDLASCDERGSDLVTYNTLS1PMEPHALVOLCGTYP	300
Db	241 HACQWALRGDADSVLSTPNSPDLASCDERGSDLVTYNTLS1PMEPHALVOLCGTYP	300
Qy	YNLTFFSSQWVLLITLITNTERRHQGFEXTFPQLPRMSSCGGR1RKXQGTNSPYRGHY	360
Db	301 YNLTFFSSQWVLLITLITNTERRHQGFEXTFPQLPRMSSCGGR1RKXQGTNSPYRGHY	360
Qy	PNPIDCTWNI1EVPNNQHVAKRFEPYLLBPGVAPACTCBKDYVE1NGEKYCGERSQFVVT	420
Db	361 PNPIDCTWNI1EVPNNQHVAKRFEPYLLBPGVAPACTCBKDYVE1NGEKYCGERSQFVVT	420
Qy	NSKKTIVRPHS00SYNDTGFLEAYL1SYSDSPCQFCTRGRC1RKE1RCDGADCTDH	480
Db	421 NSKKTIVRPHS00SYNDTGFLEAYL1SYSDSPCQFCTRGRC1RKE1RCDGADCTDH	480
Qy	SDBLNCSCDAGHQFTCKNFKCPFLFWCDSVNDCCDNDSEOGSCSPAOTFRCNSGKCLSK	540
Db	481 SDBLNCSCDAGHQFTCKNFKCPFLFWCDSVNDCCDNDSEOGSCSPAOTFRCNSGKCLSK	540
Qy	SQOCNGKDDCGSDGSDBASCPKXNVVYTCIKHYRCLNG1CLSKGNBECDDKEDGSDGDEK	600
Db	541 SQOCNGKDDCGSDGSDBASCPKXNVVYTCIKHYRCLNG1CLSKGNBECDDKEDGSDGDEK	600
Qy	DCGCGRSFTROARVVVGTDADGEGEMQWVSLHALGOGH1IGAS1ISPNNVL1SAAHCIYD	660
Db	601 DCGCGRSFTROARVVVGTDADGEGEMQWVSLHALGOGH1IGAS1ISPNNVL1SAAHCIYD	660
Qy	DRCGFRYSDFPTQWTAFLGLHDQSRSPAGVQERRLKRI1SHFPNDFTFDYDIALLEKXP	720
Db	661 DRCGFRYSDFPTQWTAFLGLHDQSRSPAGVQERRLKRI1SHFPNDFTFDYDIALLEKXP	720
Qy	AESVSWVRPLCLPDAHVFPAGKAL1WYMGHTQYGGTGALL1LQKGE1RVYNQTTGENTL	780
Db	721 AESVSWVRPLCLPDAHVFPAGKAL1WYMGHTQYGGTGALL1LQKGE1RVYNQTTGENTL	780
Qy	POQITPRMNCVGLFSLGCVDS1COGSDGGLP1SSVEADGR1FGAGV1SWGDCAGRNKPEVYT	840
Db	781 POQITPRMNCVGLFSLGCVDS1COGSDGGLP1SSVEADGR1FGAGV1SWGDCAGRNKPEVYT	840
Qy	RLPLFRDWIKENTGV 855	
Db	841 RLPLFRDWIKENTGV 855	

RESULT 5
US-10-072-012-354
; Sequence 354, Application US/10072012
; Publication No. US20040033493A1

```

/ GENERAL INFORMATION:
/ APPLICANT: Tchernyev, Vellazar
/ APPLICANT: Spytek, Kimberly
/ APPLICANT: Zethusen, Bryan
/ APPLICANT: Patturajan, Meera
/ APPLICANT: Shimkets, Richard
/ APPLICANT: Li, Li
/ APPLICANT: Ganggoli, Eshta
/ APPLICANT: Padigaru, Muralidhara
/ APPLICANT: Anderson, David W.
/ APPLICANT: Rastelli, Luca
/ APPLICANT: Miller, Charles E.
/ APPLICANT: Gerlach, Valerie
/ APPLICANT: Taupier Jr, Raymond J.
/ APPLICANT: Gusev, Vladimir Y.
/ APPLICANT: Colman, Steven D.
/ APPLICANT: Wolenc, Adam R.
/ APPLICANT: Pena, Carol B. A
/ APPLICANT: Burtak, Katarzyna
/ APPLICANT: Grosee, William M.
/ APPLICANT: Alsdbrook II, John P.
/ APPLICANT: Lepley, Denise M.
/ APPLICANT: Rieger, Daniel K.
/ APPLICANT: Burgess, Catherine B.
/ TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
/ FILE REFERENCE: 21402-258
/ CURRENT APPLICATION NUMBER: US/10/072,012
/ CURRENT FILING DATE: 2002-01-31
/ PRIOR APPLICATION NUMBER: 60/265,102
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: 60/265,514
/ PRIOR FILING DATE: 2001-01-31
/ PRIOR APPLICATION NUMBER: 60/265,517
/ PRIOR FILING DATE: 2001-01-31
/ PRIOR APPLICATION NUMBER: 60/265,412
/ PRIOR FILING DATE: 2001-01-31
/ PRIOR APPLICATION NUMBER: 60/265,395
/ PRIOR FILING DATE: 2001-01-31
/ PRIOR APPLICATION NUMBER: 60/266,406
/ PRIOR FILING DATE: 2001-02-02
/ PRIOR APPLICATION NUMBER: 60/266,767
/ PRIOR FILING DATE: 2001-02-05
/ PRIOR APPLICATION NUMBER: 60/267,057
/ PRIOR FILING DATE: 2001-02-07
/ PRIOR APPLICATION NUMBER: 60/266,975
/ PRIOR FILING DATE: 2001-02-07
/ PRIOR APPLICATION NUMBER: 60/267,459
/ PRIOR FILING DATE: 2001-02-08
/ Remaining Prior Application data removed - See File Wrapper or PALM
/ NUMBER OF SEQ ID NOS: 1391
/ SOFTWARE: Patencin Ver. 2.1
/ SEQ ID NO 354
/ LENGTH: 855
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-072-012-354

```

[illegible]

```
QY 121 KQALKLLYSGVPLGPGYHKSASVTAFASEGSAVIAAYMSEFSI PQHLYVEAEERVAEERVVM 180
Db 121 KQALKLLYSGVPLGPGYHKSASVTAFASEGSAVIAAYMSEFSI PQHLYVEAEERVAEERVVM 180
QY 181 LPPRARSLSFVYVTSVAAPPTDSKTQVORTODNSCSFGLHARGVELMRFTTPGFPDSPYPA 240
Db 181 LPPRARSLSFVYVTSVAAPPTDSKTQVORTODNSCSFGLHARGVELMRFTTPGFPDSPYPA 240
QY 241 HARCQWALRGDADSVSLTFRSPDLASCDERGSGLVYVNTLSMPEHALVOLCGTYPSPS 300
Db 241 HARCQWALRGDADSVSLTFRSPDLASCDERGSGLVYVNTLSMPEHALVOLCGTYPSPS 300
QY 301 YNLTFFHSQNVLLITLITNTERRHHPGEATFPQLPRMSCGRLRKAQGTENSPYPGAY 360
Db 301 YNLTFFHSQNVLLITLITNTERRHHPGEATFPQLPRMSCGRLRKAQGTENSPYPGAY 360
QY 361 PPNIDCTWNIIEVNNQHVAVRFKFFYLLEPGVAGTCKPYVEINGEKTCGERSQFVYTS 420
Db 361 PPNIDCTWNIIEVNNQHVAVRFKFFYLLEPGVAGTCKPYVEINGEKTCGERSQFVYTS 420
QY 421 NSNKITVRPHSDOSYTDYTGFLAEYLSYSDSDPCGQFTCRGRCIRKELRCDDMADCTDH 480
Db 421 NSNKITVRPHSDOSYTDYTGFLAEYLSYSDSDPCGQFTCRGRCIRKELRCDDMADCTDH 480
QY 481 SDEINCSGADGHOFTCKNKECKPLFWYCDSDVNDGSDNSDBQSCSPAQTRFCSNGKCLSK 540
Db 481 SDEINCSGADGHOFTCKNKECKPLFWYCDSDVNDGSDNSDBQSCSPAQTRFCSNGKCLSK 540
QY 541 SQQNGDGDGSDSDASCKRANVYVTTKXTYRCLNLCISKNPEDCDGSDSDSK 600
Db 541 SQQNGDGDGSDSDASCKRANVYVTTKXTYRCLNLCISKNPEDCDGSDSDSK 600
QY 601 DCCGGLSFTROARVVGATDADEGEMPOYSLHLAGGSHICGASLISPMNLVSAHICYID 660
Db 601 DCCGGLSFTROARVVGATDADEGEMPOYSLHLAGGSHICGASLISPMNLVSAHICYID 660
QY 661 DRGFRYSPTQWTAFLGLHDQSORAPGVQERLRKRIISHPFNDTFDYDIALLELEKP 720
Db 661 DRGFRYSPTQWTAFLGLHDQSORAPGVQERLRKRIISHPFNDTFDYDIALLELEKP 720
QY 721 AEYSSMWRTICLPDASHVFPAGKAIWTTGNGHTOYGGTGALLIOKGIRIYNQTCNLL 780
Db 721 AEYSSMWRTICLPDASHVFPAGKAIWTTGNGHTOYGGTGALLIOKGIRIYNQTCNLL 780
QY 781 PQQITPRMCCVGLSGVDSCQSDSGPLSSVEADGRIFGAGVYVSWDGCAGRRKPGVYT 840
Db 781 PQQITPRMCCVGLSGVDSCQSDSGPLSSVEADGRIFGAGVYVSWDGCAGRRKPGVYT 840
QY 841 RLPLFRDWIKENTGV 855
Db 841 RLPLFRDWIKENTGV 855

RESULT 6
US-10-072-012-420
; Sequence 420, Application US/10072012
; Publication No. US2004003493A1
; GENERAL INFORMATION:
; APPLICANT: Tchernev, Velizar
; APPLICANT: Spytak, Kimberly
; APPLICANT: Zehnusen, Bryan
; APPLICANT: Paturajan, Meera
; APPLICANT: Shinkets, Richard
; APPLICANT: Li, Li
; APPLICANT: Gangoli, Esha
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Anderson, David W.
; APPLICANT: Rastelli, Luca
; APPLICANT: Miller, Charles B.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Tsaplier, Jr, Raymond J.
; APPLICANT: Gusev, Vladimyr Y.
; APPLICANT: Colman, Steven D.
```

```
APPLICANT: Molenc, Adam R.
APPLICANT: Pena, Carol E. A
APPLICANT: Furtak, Katarzyna
APPLICANT: Grose, William M.
APPLICANT: Alsodrook II, John P.
APPLICANT: Lepley, Denise M.
APPLICANT: Rieger, Daniel K.
APPLICANT: Burgess, Catherine E.
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
FILE REFERENCE: 21402-258
CURRENT APPLICATION NUMBER: US/10/072, 012
CURRENT FILING DATE: 2002-01-31
PRIOR APPLICATION NUMBER: 60/265, 102
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: 60/265, 514
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265, 517
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265, 412
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265, 395
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/266, 406
PRIOR FILING DATE: 2001-02-02
PRIOR APPLICATION NUMBER: 60/266, 767
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: 60/267, 057
PRIOR FILING DATE: 2001-02-07
PRIOR APPLICATION NUMBER: 60/266, 975
PRIOR FILING DATE: 2001-02-07
PRIOR APPLICATION NUMBER: 60/267, 459
PRIOR FILING DATE: 2001-02-08
Remainder Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 1391
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 420
LENGTH: 855
TYPE: PRT
ORGANISM: Homo sapiens
US-10-072-012-420

Query Match 99.4%; Score 4655; DB 15; Length 855;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 MSDBRARKGGGKPGDAGIKYNSRHEKXNGLEBGEVFPVNNVKKYERKGPQRWVLLA 60
Db 1 MSDBRARKGGGKPGDAGIKYNSRHEKXNGLEBGEVFPVNNVKKYERKGPQRWVLLA 60
QY 61 VILIGLLVILGIGFLVMHLQYRDVRYQYKNGYMRITNENFVDAYENSNSTEFVSLASRY 120
Db 61 VILIGLLVILGIGFLVMHLQYRDVRYQYKNGYMRITNENFVDAYENSNSTEFVSLASRY 120
QY 121 KQALKLLYSGVPLGPGYHKSASVTAFASEGSAVIAAYMSEFSI PQHLYVEAEERVAEERVVM 180
Db 121 KQALKLLYSGVPLGPGYHKSASVTAFASEGSAVIAAYMSEFSI PQHLYVEAEERVAEERVVM 180
QY 181 LPPRARSLSFVYVTSVAAPPTDSKTQVORTODNSCSFGLHARGVELMRFTTPGFPDSPYPA 240
Db 181 LPPRARSLSFVYVTSVAAPPTDSKTQVORTODNSCSFGLHARGVELMRFTTPGFPDSPYPA 240
QY 241 HARCQWALRGDADSVSLTFRSPDLASCDERGSGLVYVNTLSMPEHALVOLCGTYPSPS 300
Db 241 HARCQWALRGDADSVSLTFRSPDLASCDERGSGLVYVNTLSMPEHALVOLCGTYPSPS 300
QY 301 YNLTFFHSQNVLLITLITNTERRHHPGEATFPQLPRMSCGRLRKAQGTENSPYPGAY 360
Db 301 YNLTFFHSQNVLLITLITNTERRHHPGEATFPQLPRMSCGRLRKAQGTENSPYPGAY 360
QY 361 PPNIDCTWNIIEVNNQHVAVRFKFFYLLEPGVAGTCKPYVEINGEKTCGERSQFVYTS 420
Db 361 PPNIDCTWNIIEVNNQHVAVRFKFFYLLEPGVAGTCKPYVEINGEKTCGERSQFVYTS 420
```

QY 421 NSNKITVRFHSDQSYDTDTGFLAEYLSDSDPCPGQFTCRGTGRCKIRKELRCGDNADCTDH 480
Db 421 NSNKITVRFHSDQSYDTDTGFLAEYLSDSDPCPGQFTCRGTGRCKIRKELRCGDNADCTDH 480
QY 481 SDELNCSGDAGHQTCKNKKFKPLFWVCDVNDGDNDEGQCCGPAQTFRCSNGKLSK 540
Db 481 SDELNCSGDAGHQTCKNKKFKPLFWVCDVNDGDNDEGQCCGPAQTFRCSNGKLSK 540
QY 541 SQQNGKDDCGDSDGDEASCPVNVVCTCKHYRCLNGLCLSKNGECCGKEDSDGSEK 600
Db 541 SQQNGKDDCGDSDGDEASCPVNVVCTCKHYRCLNGLCLSKNGECCGKEDSDGSEK 600
QY 601 DDCGLSFTROARVVGSTDADEGEMPVQVSLHALGGHICGASLISPNMLVSAHCTID 660
Db 601 DDCGLSFTROARVVGSTDADEGEMPVQVSLHALGGHICGASLISPNMLVSAHCTID 660
QY 661 DRGFRYSPTQWTAFLGLHDSQRSAPGVQERRLRKIIISHPFNDFTFDYDIALLELEK 720
Db 661 DRGFRYSPTQWTAFLGLHDSQRSAPGVQERRLRKIIISHPFNDFTFDYDIALLELEK 720
QY 721 AEYSMWPRICLPDASHVFPAGKAIWVGWGTGCGTGGALILQKGEIRVINOCTCEML 780
Db 721 AEYSMWPRICLPDASHVFPAGKAIWVGWGTGCGTGGALILQKGEIRVINOCTCEML 780
QY 781 PQQITPRMVCVFLSGVDSQCGSDGGLSSVEADGRIFGAGVSWGDCAGRNKPGVYT 840
Db 781 PQQITPRMVCVFLSGVDSQCGSDGGLSSVEADGRIFGAGVSWGDCAGRNKPGVYT 840
QY 841 RLPLFRDWIKENTGV 855
Db 841 RLPLFRDWIKENTGV 855

RESULT 7

US-10-037-417-132
Sequence 132, Application US/10037417
Publication No. US20040052806A1
GENERAL INFORMATION:
APPLICANT: Kekuda, Rameeh
APPLICANT: Alsobrook II, John P
APPLICANT: Tcherenev, Veilizar T
APPLICANT: Liu, Xiaohong
APPLICANT: Spytek, Kimberly A
APPLICANT: Patturajan, Meera
APPLICANT: Groose, William M
APPLICANT: Lepley, Denise M
APPLICANT: Burgess, Catherine E
APPLICANT: Li, Li
APPLICANT: Gorman, Linda
APPLICANT: Edinger, Shlomit R
APPLICANT: Scioire, Paul
APPLICANT: Ellerman, Karen
APPLICANT: Malyankar, Uriel M
APPLICANT: Rothenberg, Mark
APPLICANT: Stone, David J
APPLICANT: Boldog, Ferenc L
APPLICANT: Guo, Xiaojia
APPLICANT: Shenoy, Suresh G
APPLICANT: Anderson, David W
APPLICANT: Padigaru, Murallidhara
APPLICANT: Taupier Jr, Raymond J
APPLICANT: Miller, Charles E
APPLICANT: Eileen, Andrew J
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
FILE REFERENCE: 21402-235
CURRENT APPLICATION NUMBER: US/10/037,417
CURRENT FILING DATE: 2002-09-20
PRIOR APPLICATION NUMBER: 60/260,018
PRIOR FILING DATE: 2001-01-05
PRIOR APPLICATION NUMBER: 60/260,360
PRIOR FILING DATE: 2001-01-08
PRIOR APPLICATION NUMBER: 60/272,411

PRIOR FILING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: 60/272,817
PRIOR FILING DATE: 2001-03-02
PRIOR APPLICATION NUMBER: 60/291,186
PRIOR FILING DATE: 2001-05-15
PRIOR APPLICATION NUMBER: 60/303,231
PRIOR FILING DATE: 2001-07-05
PRIOR APPLICATION NUMBER: 60/305,060
PRIOR FILING DATE: 2001-07-12
PRIOR APPLICATION NUMBER: 60/318,405
PRIOR FILING DATE: 2001-09-10
PRIOR APPLICATION NUMBER: 60/318,700
PRIOR FILING DATE: 2001-09-12
NUMBER OF SEQ ID NOS: 227
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 132
LENGTH: 855
TYPE: PRT
ORGANISM: Homo sapiens
US-10-037-417-132

Query Match 99.4%; Score 4655; DB 15; Length 855;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 MGSRRARKGGGPGDPGAGLKYNRSRHEKNGLEGEVFLPVNNVKKYKKGPGRWVLLAA 60
Db 1 MGSRRARKGGGPGDPGAGLKYNRSRHEKNGLEGEVFLPVNNVKKYKKGPGRWVLLAA 60
QY 61 VLIGLLVLLIGLGLVHMLQYRDVRYQKXNGYMRITNENFVADAYENSSTFVSLASKV 120
Db 61 VLIGLLVLLIGLGLVHMLQYRDVRYQKXNGYMRITNENFVADAYENSSTFVSLASKV 120
QY 121 KDALKLIYSGVPLGPRHKSATVTAPEGSVIAIYNNSEFSIPQULVEARVMAEERVVM 180
Db 121 KDALKLIYSGVPLGPRHKSATVTAPEGSVIAIYNNSEFSIPQULVEARVMAEERVVM 180
QY 121 KDALKLIYSGVPLGPRHKSATVTAPEGSVIAIYNNSEFSIPQULVEARVMAEERVVM 180
Db 121 KDALKLIYSGVPLGPRHKSATVTAPEGSVIAIYNNSEFSIPQULVEARVMAEERVVM 180
QY 181 LPPRARSLSKFVTVSVVAFPTDSKTVOPTDONSCEFGLHARGVLMFTTGGPDSRYPA 240
Db 181 LPPRARSLSKFVTVSVVAFPTDSKTVOPTDONSCEFGLHARGVLMFTTGGPDSRYPA 240
QY 241 HARCOMALRGDADSVLSLTFRSPDLASCDBRGSDLVTYNTLSPEMHALVOLCGTYP 300
Db 241 HARCOMALRGDADSVLSLTFRSPDLASCDBRGSDLVTYNTLSPEMHALVOLCGTYP 300
QY 241 HARCOMALRGDADSVLSLTFRSPDLASCDBRGSDLVTYNTLSPEMHALVOLCGTYP 300
Db 241 HARCOMALRGDADSVLSLTFRSPDLASCDBRGSDLVTYNTLSPEMHALVOLCGTYP 300
QY 301 YNLTFFSSQVNLITLITNTERRRHGFBEATFPQLPKMSGCCGRLRKAQGTNPSYTYGHY 360
Db 301 YNLTFFSSQVNLITLITNTERRRHGFBEATFPQLPKMSGCCGRLRKAQGTNPSYTYGHY 360
QY 361 PPNIDCTWNIIEVPPNQHVKVRFKFFYLEBPGVAGTCPKDYVEINERKYGERSQFVYTS 420
Db 361 PPNIDCTWNIIEVPPNQHVKVRFKFFYLEBPGVAGTCPKDYVEINERKYGERSQFVYTS 420
QY 421 NSNKITVRFHSDQSYDTDTGFLAEYLSDSDPCPGQFTCRGTGRCKIRKELRCGDNADCTDH 480
Db 421 NSNKITVRFHSDQSYDTDTGFLAEYLSDSDPCPGQFTCRGTGRCKIRKELRCGDNADCTDH 480
QY 481 SDELNCSGDAGHQTCKNKKFKPLFWVCDVNDGDNDEGQCCGPAQTFRCSNGKLSK 540
Db 481 SDELNCSGDAGHQTCKNKKFKPLFWVCDVNDGDNDEGQCCGPAQTFRCSNGKLSK 540
QY 541 SQQNGKDDCGDSDGDEASCPVNVVCTCKHYRCLNGLCLSKNGECCGKEDSDGSEK 600
Db 541 SQQNGKDDCGDSDGDEASCPVNVVCTCKHYRCLNGLCLSKNGECCGKEDSDGSEK 600
QY 601 DDCGLSFTROARVVGSTDADEGEMPVQVSLHALGGHICGASLISPNMLVSAHCTID 660
Db 601 DDCGLSFTROARVVGSTDADEGEMPVQVSLHALGGHICGASLISPNMLVSAHCTID 660
QY 661 DRGFRYSPTQWTAFLGLHDSQRSAPGVQERRLRKIIISHPFNDFTFDYDIALLELEK 720
Db 661 DRGFRYSPTQWTAFLGLHDSQRSAPGVQERRLRKIIISHPFNDFTFDYDIALLELEK 720

```
QY 721 AEYSWVRPCLPDASHVFPAGKAIWVTGNGHTQYGGTALLIQKEIRVINOCTCENLL 780
DB 721 AEYSWVRPCLPDASHVFPAGKAIWVTGNGHTQYGGTALLIQKEIRVINOCTCENLL 780
QY 781 PQQITPRMNCVGLSGVDSCQDSGGPLSVADGRIFGAGVVSNGDGCAGRNKPGVYT 840
DB 781 PQQITPRMNCVGLSGVDSCQDSGGPLSVADGRIFGAGVVSNGDGCAGRNKPGVYT 840
QY 841 RLPLFRDWIKENTGV 855
DB 841 RLPLFRDWIKENTGV 855

RESULT 8
US-09-776-191-2
; Sequence 2, Application US/09776191
; Publication No. US20030119168A1
; GENERAL INFORMATION:
; APPLICANT: Edwin L. Madison
; APPLICANT: Edgar O. Ong
; APPLICANT: Jiumn-Chern Yeh
; APPLICANT: Corvax International, Inc.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; TITLE OF INVENTION: TRANSMEMBRANE SERINE PROTEASES, THE ENCODED PROTEINS AND
; FILE REFERENCE: 24745-1607
; CURRENT APPLICATION NUMBER: US/09/776,191
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/213,124
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 60/234,840
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 60/179,982
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: 60/183,542
; PRIOR FILING DATE: 2000-02-18
; PRIOR APPLICATION NUMBER: 09/657,968
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: PASTESEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-776-191-2

Query Match 99.4%; Score 4653; DB 10; Length 855;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 MGSDBARKGGGGKDPFGAGLKYNSRHEKVNGLSEGVFPLPVNNVKKYERKHGPGRWVLLAA 60
DB 1 MGSDBARKGGGGKDPFGAGLKYNSRHEKVNGLSEGVFPLPVNNVKKYERKHGPGRWVLLAA 60
QY 61 VLIGLLLVLLIGLIGFLVWHLQYRDVAVQKXNGYMRITNENFVDAYENSNSTEFVSLASKV 120
DB 61 VLIGLLLVLLIGLIGFLVWHLQYRDVAVQKXNGYMRITNENFVDAYENSNSTEFVSLASKV 120
QY 121 KDALKLISGVPLGPGYHKSAVTAPESEGVIAIYMWSEFSIPQHLVEABERVAEERVVM 180
DB 121 KDALKLISGVPLGPGYHKSAVTAPESEGVIAIYMWSEFSIPQHLVEABERVAEERVVM 180
QY 181 LPPPARSLKSFVVVSVAAPPTDSKTVORTODNSCSFGIHHARGVLMFTTTPGFPDSYPYA 240
DB 181 LPPPARSLKSFVVVSVAAPPTDSKTVORTODNSCSFGIHHARGVLMFTTTPGFPDSYPYA 240
QY 241 HARCQMLRGDADSVLSLTFERSFDLASCDERGSDLVTVYNTLSPEBHALVOLCGTYPSPS 300
DB 241 HARCQMLRGDADSVLSLTFERSFDLASCDERGSDLVTVYNTLSPEBHALVOLCGTYPSPS 300
QY 301 YNLTFFSSQNVLLTTLTNTERRRHPGFPAFFPOLPRMSSCGRLRKAQGTFSNPPYFGHY 360
DB 301 YNLTFFSSQNVLLTTLTNTERRRHPGFPAFFPOLPRMSSCGRLRKAQGTFSNPPYFGHY 360
```

```
QY 361 PENIDCTWNIIEVNNQHVKRFKFFYLEPBGVPACTCPKDYVEINCKTCGERSQEVVTS 420
DB 361 PENIDCTWNIIEVNNQHVKRFKFFYLEPBGVPACTCPKDYVEINCKTCGERSQEVVTS 420
QY 421 NSNKTIVRRHSDDSYDTGFLAEYLSYSSDPCEQGFCTGTCIRKEIRCDGMADCTTH 480
DB 421 NSNKTIVRRHSDDSYDTGFLAEYLSYSSDPCEQGFCTGTCIRKEIRCDGMADCTTH 480
QY 481 SDELNCSDAAGHOFCTCKNFKCKPLFWVCDSDVNDGNSDEQSCSPAOTFRCSNGKLSK 540
DB 481 SDELNCSDAAGHOFCTCKNFKCKPLFWVCDSDVNDGNSDEQSCSPAOTFRCSNGKLSK 540
QY 541 SQQCNKGDQDGSDEASCPKXNVVYCTKHTYRCNLGLCLSKNPECDGKEDCSGSDSK 600
DB 541 SQQCNKGDQDGSDEASCPKXNVVYCTKHTYRCNLGLCLSKNPECDGKEDCSGSDSK 600
QY 601 DCCGGLRSFTFRQARVVGCTDADGEMPMQVSLHALQGHICGASLISPMVLVSAACIYD 660
DB 601 DCCGGLRSFTFRQARVVGCTDADGEMPMQVSLHALQGHICGASLISPMVLVSAACIYD 660
QY 661 DRGFRYSDPQTQWTAFLGLHDQSORAPGVQERRLKRIISHPPNDFTFDYDIALLEKXP 720
DB 661 DRGFRYSDPQTQWTAFLGLHDQSORAPGVQERRLKRIISHPPNDFTFDYDIALLEKXP 720
QY 721 AEYSWVRPCLPDASHVFPAGKAIWVTGNGHTQYGGTALLIQKEIRVINOCTCENLL 780
DB 721 AEYSWVRPCLPDASHVFPAGKAIWVTGNGHTQYGGTALLIQKEIRVINOCTCENLL 780
QY 781 PQQITPRMNCVGLSGVDSCQDSGGPLSVADGRIFGAGVVSNGDGCAGRNKPGVYT 840
DB 781 PQQITPRMNCVGLSGVDSCQDSGGPLSVADGRIFGAGVVSNGDGCAGRNKPGVYT 840
QY 841 RLPLFRDWIKENTGV 855
DB 841 RLPLFRDWIKENTGV 855

RESULT 9
US-10-099-700A-2
; Sequence 2, Application US/10099700A
; Publication No. US20030008372A1
; GENERAL INFORMATION:
; APPLICANT: Edwin L. Madison
; APPLICANT: Edgar O. Ong
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING TRANSMEMBRANE SERINE PROTEASE 7, 1
; TITLE OF INVENTION: ENCODED POLYPEPTIDES AND METHODS BASED THEREON
; FILE REFERENCE: 24745-1613
; CURRENT APPLICATION NUMBER: US/10/099,700A
; PRIOR FILING DATE: 2002-05-24
; PRIOR APPLICATION NUMBER: 60/275,592
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PASTESEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-099-700A-2

Query Match 99.4%; Score 4653; DB 14; Length 855;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 MGSDBARKGGGGKDPFGAGLKYNSRHEKVNGLSEGVFPLPVNNVKKYERKHGPGRWVLLAA 60
DB 1 MGSDBARKGGGGKDPFGAGLKYNSRHEKVNGLSEGVFPLPVNNVKKYERKHGPGRWVLLAA 60
QY 61 VLIGLLLVLLIGLIGFLVWHLQYRDVAVQKXNGYMRITNENFVDAYENSNSTEFVSLASKV 120
DB 61 VLIGLLLVLLIGLIGFLVWHLQYRDVAVQKXNGYMRITNENFVDAYENSNSTEFVSLASKV 120
QY 121 KDALKLISGVPLGPGYHKSAVTAPESEGVIAIYMWSEFSIPQHLVEABERVAEERVVM 180
DB 121 KDALKLISGVPLGPGYHKSAVTAPESEGVIAIYMWSEFSIPQHLVEABERVAEERVVM 180
```

```
Db 121 KDALKLYSGVPLGPHKESAVTAFAEGSVIAYYMSFSIPQHLVEARVMAEERVVM 180
Qy 181 LPPRARSLSKSFVVTSVVAFPTDSKTVOQTQDNCSFGIHAARGVELMFTTGGPPDSPPYA 240
Db 181 LPPRARSLSKSFVVTSVVAFPTDSKTVOQTQDNCSFGIHAARGVELMFTTGGPPDSPPYA 240
Qy 241 HARCQMLRGDADSVLSTFRSPDLASCDERGSIDLVTYNTLSPEMPEHALVOLCGTYP 300
Db 241 HARCQMLRGDADSVLSTFRSPDLASCDERGSIDLVTYNTLSPEMPEHALVOLCGTYP 300
Qy 301 YNLTFHSSQNVLLITLITNTERRHGPEATFPOLPRMSSCGGRRLRKAQGTFSPPYPGHY 360
Db 301 YNLTFHSSQNVLLITLITNTERRHGPEATFPOLPRMSSCGGRRLRKAQGTFSPPYPGHY 360
Qy 361 PPNIIDCTWNIIEVPNNQHVKVPFKPFYLLBPGVPAGTCKPDYVEINGEKYCGERSQFVVT 420
Db 361 PPNIIDCTWNIIEVPNNQHVKVPFKPFYLLBPGVPAGTCKPDYVEINGEKYCGERSQFVVT 420
Qy 421 NSNKITVRFHSDOSTYDTGFLAEYLSYDSDPCPGQFTCTGRGCIKRELCDGMADCTDH 480
Db 421 NSNKITVRFHSDOSTYDTGFLAEYLSYDSDPCPGQFTCTGRGCIKRELCDGMADCTDH 480
Qy 481 SDELINCSGDAGHQTCKNKFCKPLFWVCDVNDCGDSDQSCSPAQTFRCNSGKCLSK 540
Db 481 SDELINCSGDAGHQTCKNKFCKPLFWVCDVNDCGDSDQSCSPAQTFRCNSGKCLSK 540
Qy 541 SQQNGKDDCGDSDGDEASCPKVVNVVCTKHTYRCLNGLCLSKNPECDGKEDCSGSDSEK 600
Db 541 SQQNGKDDCGDSDGDEASCPKVVNVVCTKHTYRCLNGLCLSKNPECDGKEDCSGSDSEK 600
Qy 601 DCDGGLSFTROARVVGTTADDEGEWPMQVSLHALGGHICGASLISPNMLVSAHCTID 660
Db 601 DCDGGLSFTROARVVGTTADDEGEWPMQVSLHALGGHICGASLISPNMLVSAHCTID 660
Qy 661 DRGFRYSPTQMTAFGLHDOSORSAPGVQERRLKRIISHPFNDFTFDYDIALLELEK 720
Db 661 DRGFRYSPTQMTAFGLHDOSORSAPGVQERRLKRIISHPFNDFTFDYDIALLELEK 720
Qy 721 AEYSMWVRPICLPDASHVFPAGKAIWVGWGHYGGTGALLILOGEIRIVINQTTCEML 780
Db 721 AEYSMWVRPICLPDASHVFPAGKAIWVGWGHYGGTGALLILOGEIRIVINQTTCEML 780
Qy 781 PQQITPRMVCVGLSGVDSCQSDSGPLSSVEADGRIFGAGVSWKGGCGARKKPGVYT 840
Db 781 PQQITPRMVCVGLSGVDSCQSDSGPLSSVEADGRIFGAGVSWKGGCGARKKPGVYT 840
Qy 841 RLPLFRDIKENTGV 855
Db 841 RLPLFRDIKENTGV 855

RESULT 10
US-10-190-030B-2
; Sequence 2, Application US/10190030B
; Publication No. US20030134298A1
; GENERAL INFORMATION:
; APPLICANT: Madison, Edwin
; APPLICANT: Ong, Edgar
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING A
; TITLE OF INVENTION: TRANSMEMBRANE SERINE PROTEASE 20, THE ENCODED POLYPEPTIDES AND
; FILE REFERENCE: 24745-1618
; CURRENT APPLICATION NUMBER: US/10/190, 030B
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-190-030B-2
```

```
Query Match 99.4%; Score 4653; DB 14; Length 855;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 MGSDBRARKGGGGKDFGAGLKYNSRHEKVNGLBEGVEFLPVNNVKKYKXGPGRMVLLAA 60
Db 1 MGSDBRARKGGGGKDFGAGLKYNSRHEKVNGLBEGVEFLPVNNVKKYKXGPGRMVLLAA 60
Qy 61 VLIGLLLVLLIGIFLVNHLQYRDYRVQKXNGYKRIINENPVDAEENSSTEFVSLASKV 120
Db 61 VLIGLLLVLLIGIFLVNHLQYRDYRVQKXNGYKRIINENPVDAEENSSTEFVSLASKV 120
Qy 121 KDALKLYSGVPLGPHKESAVTAFAEGSVIAYYMSFSIPQHLVEARVMAEERVVM 180
Db 121 KDALKLYSGVPLGPHKESAVTAFAEGSVIAYYMSFSIPQHLVEARVMAEERVVM 180
Qy 181 LPPRARSLSKSFVVTSVVAFPTDSKTVOQTQDNCSFGIHAARGVELMFTTGGPPDSPPYA 240
Db 181 LPPRARSLSKSFVVTSVVAFPTDSKTVOQTQDNCSFGIHAARGVELMFTTGGPPDSPPYA 240
Qy 241 HARCQMLRGDADSVLSTFRSPDLASCDERGSIDLVTYNTLSPEMPEHALVOLCGTYP 300
Db 241 HARCQMLRGDADSVLSTFRSPDLASCDERGSIDLVTYNTLSPEMPEHALVOLCGTYP 300
Qy 301 YNLTFHSSQNVLLITLITNTERRHGPEATFPOLPRMSSCGGRRLRKAQGTFSPPYPGHY 360
Db 301 YNLTFHSSQNVLLITLITNTERRHGPEATFPOLPRMSSCGGRRLRKAQGTFSPPYPGHY 360
Qy 361 PPNIIDCTWNIIEVPNNQHVKVPFKPFYLLBPGVPAGTCKPDYVEINGEKYCGERSQFVVT 420
Db 361 PPNIIDCTWNIIEVPNNQHVKVPFKPFYLLBPGVPAGTCKPDYVEINGEKYCGERSQFVVT 420
Qy 421 NSNKITVRFHSDOSTYDTGFLAEYLSYDSDPCPGQFTCTGRGCIKRELCDGMADCTDH 480
Db 421 NSNKITVRFHSDOSTYDTGFLAEYLSYDSDPCPGQFTCTGRGCIKRELCDGMADCTDH 480
Qy 481 SDELINCSGDAGHQTCKNKFCKPLFWVCDVNDCGDSDQSCSPAQTFRCNSGKCLSK 540
Db 481 SDELINCSGDAGHQTCKNKFCKPLFWVCDVNDCGDSDQSCSPAQTFRCNSGKCLSK 540
Qy 541 SQQNGKDDCGDSDGDEASCPKVVNVVCTKHTYRCLNGLCLSKNPECDGKEDCSGSDSEK 600
Db 541 SQQNGKDDCGDSDGDEASCPKVVNVVCTKHTYRCLNGLCLSKNPECDGKEDCSGSDSEK 600
Qy 601 DCDGGLSFTROARVVGTTADDEGEWPMQVSLHALGGHICGASLISPNMLVSAHCTID 660
Db 601 DCDGGLSFTROARVVGTTADDEGEWPMQVSLHALGGHICGASLISPNMLVSAHCTID 660
Qy 661 DRGFRYSPTQMTAFGLHDOSORSAPGVQERRLKRIISHPFNDFTFDYDIALLELEK 720
Db 661 DRGFRYSPTQMTAFGLHDOSORSAPGVQERRLKRIISHPFNDFTFDYDIALLELEK 720
Qy 721 AEYSMWVRPICLPDASHVFPAGKAIWVGWGHYGGTGALLILOGEIRIVINQTTCEML 780
Db 721 AEYSMWVRPICLPDASHVFPAGKAIWVGWGHYGGTGALLILOGEIRIVINQTTCEML 780
Qy 781 PQQITPRMVCVGLSGVDSCQSDSGPLSSVEADGRIFGAGVSWKGGCGARKKPGVYT 840
Db 781 PQQITPRMVCVGLSGVDSCQSDSGPLSSVEADGRIFGAGVSWKGGCGARKKPGVYT 840
Qy 841 RLPLFRDIKENTGV 855
Db 841 RLPLFRDIKENTGV 855

RESULT 11
US-10-302-840A-2
; Sequence 2, Application US/10302840A
; Publication No. US20030134794A1
; GENERAL INFORMATION:
; APPLICANT: Madison, Edwin L.
; APPLICANT: Ong, Edgar O.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING SERINE PROTEASE CVSP17, THE ENCODED
```

```
; TITLE OF INVENTION: POLYPEPTIDES AND METHODS BASED THEREON
; FILE REFERENCE: 24745-1622
; CURRENT APPLICATION NUMBER: US/10/302,840A
; CURRENT FILING DATE: 2003-01-24
; PRIOR APPLICATION NUMBER: 60/332,015
; PRIOR FILING DATE: 2001-11-20
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-302-840A-2

Query Match      99.4%; Score 4653; DB 14; Length 855;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 MGSBRARKGGGGPKDFGAGLKTNSRHEKNGLEGEVEFLPVNNVKKYKKGPGRMVVLAA 60
DB      1 MGSBRARKGGGGPKDFGAGLKTNSRHEKNGLEGEVEFLPVNNVKKYKKGPGRMVVLAA 60
QY      61 VLIIGLLVLIIGIFLVWHLQYRDVAVOKYKNGYMRITNENPVDAYENSSTEFVSLASKV 120
DB      61 VLIIGLLVLIIGIFLVWHLQYRDVAVOKYKNGYMRITNENPVDAYENSSTEFVSLASKV 120
QY      121 KDALKLLYSGVPLGPHKESAVTAFSEGSVIAYYSEFSIPQHLVEAEERVAEERVVM 180
DB      121 KDALKLLYSGVPLGPHKESAVTAFSEGSVIAYYSEFSIPQHLVEAEERVAEERVVM 180
QY      121 KDALKLLYSGVPLGPHKESAVTAFSEGSVIAYYSEFSIPQHLVEAEERVAEERVVM 180
DB      121 KDALKLLYSGVPLGPHKESAVTAFSEGSVIAYYSEFSIPQHLVEAEERVAEERVVM 180
QY      181 LPPRARSLSKSFVVTSSVAAPFTDSKTVOQTODNSCSFGLHARGVELMRFTTGPFPDSYPYA 240
DB      181 LPPRARSLSKSFVVTSSVAAPFTDSKTVOQTODNSCSFGLHARGVELMRFTTGPFPDSYPYA 240
QY      181 LPPRARSLSKSFVVTSSVAAPFTDSKTVOQTODNSCSFGLHARGVELMRFTTGPFPDSYPYA 240
DB      181 LPPRARSLSKSFVVTSSVAAPFTDSKTVOQTODNSCSFGLHARGVELMRFTTGPFPDSYPYA 240
QY      241 HARCOMALRGDADSVLSLTFRSFDLASCDERGSDLVTVYNTLSPMEPHALVOLCGTYPSPS 300
DB      241 HARCOMALRGDADSVLSLTFRSFDLASCDERGSDLVTVYNTLSPMEPHALVOLCGTYPSPS 300
QY      301 YNLTFFSSQNVLLITLITNTERRHGPGEATFFQLPRMSSCGGLRKAQGTFFNSPYYPGHY 360
DB      301 YNLTFFSSQNVLLITLITNTERRHGPGEATFFQLPRMSSCGGLRKAQGTFFNSPYYPGHY 360
QY      361 PPNIDCTWNIEVNNQHVKKRFFKFFYLLEPGVPAGTCPKDYVEINGEKYCGERSQFVYVS 420
DB      361 PPNIDCTWNIEVNNQHVKKRFFKFFYLLEPGVPAGTCPKDYVEINGEKYCGERSQFVYVS 420
QY      421 NSNKITVRFFSDOSYTDITGFLAEYLSYDSDPCPGQFTCTGTGCIKRELRCDMACTDH 480
DB      421 NSNKITVRFFSDOSYTDITGFLAEYLSYDSDPCPGQFTCTGTGCIKRELRCDMACTDH 480
QY      481 SDELINCSADAGHOFCTCKNFKCKPLFWVCDSVNDGSDSDEQSCSPAQTFRCSNGKCLSK 540
DB      481 SDELINCSADAGHOFCTCKNFKCKPLFWVCDSVNDGSDSDEQSCSPAQTFRCSNGKCLSK 540
QY      541 SQQCNMGDDCGSDSDEASCPKVVNVVTCTKTYRCLNGLCLSKNPPCEDGKEDCSGSDSK 600
DB      541 SQQCNMGDDCGSDSDEASCPKVVNVVTCTKTYRCLNGLCLSKNPPCEDGKEDCSGSDSK 600
QY      601 DCDGGLSFTROARVVGCTDADEGEWPMOYSLHALGQGHICGASLISPMNLVSAAHCIYD 660
DB      601 DCDGGLSFTROARVVGCTDADEGEWPMOYSLHALGQGHICGASLISPMNLVSAAHCIYD 660
QY      661 DRGRFYSDPQTOWTAFGLIHDSQSRAPGVDERRLKRIISHPFNDFTFYDIALLELEKP 720
DB      661 DRGRFYSDPQTOWTAFGLIHDSQSRAPGVDERRLKRIISHPFNDFTFYDIALLELEKP 720
QY      721 AEVSSMRPITCLPASHVFPAGKAIWYTGWGHITQYGGTGAIILOKGEIRVYNQTTGNNL 780
DB      721 AEVSSMRPITCLPASHVFPAGKAIWYTGWGHITQYGGTGAIILOKGEIRVYNQTTGNNL 780
QY      781 PQQITPRMMCVGFLSGGVDSQSGGPGPLSSVEADGRIFGAGVYVSWGDGCGAGRRKPGVYT 840
DB      781 PQQITPRMMCVGFLSGGVDSQSGGPGPLSSVEADGRIFGAGVYVSWGDGCGAGRRKPGVYT 840
```

```
QY      841 RLPLFRDWIKENTGV 855
DB      841 RLPLFRDWIKENTGV 855

RESULT 12
US-10-267-219-2
; Sequence 2, Application US/10267219
; Publication No. US20030143219A1
; GENERAL INFORMATION:
; APPLICANT: Madison, Edwin
; APPLICANT: Jem, Jium-Chern
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING A TRANSMEMBRANE SERINE PROTEASE 2;
; FILE REFERENCE: 24745-1621
; CURRENT APPLICATION NUMBER: US/10/267,219
; CURRENT FILING DATE: 2002-10-08
; PRIOR APPLICATION NUMBER: 60/328,530
; PRIOR FILING DATE: 09-OCT-2001
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-267-219-2

Query Match      99.4%; Score 4653; DB 14; Length 855;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 MGSBRARKGGGGPKDFGAGLKTNSRHEKNGLEGEVEFLPVNNVKKYKKGPGRMVVLAA 60
DB      1 MGSBRARKGGGGPKDFGAGLKTNSRHEKNGLEGEVEFLPVNNVKKYKKGPGRMVVLAA 60
QY      61 VLIIGLLVLIIGIFLVWHLQYRDVAVOKYKNGYMRITNENPVDAYENSSTEFVSLASKV 120
DB      61 VLIIGLLVLIIGIFLVWHLQYRDVAVOKYKNGYMRITNENPVDAYENSSTEFVSLASKV 120
QY      121 KDALKLLYSGVPLGPHKESAVTAFSEGSVIAYYSEFSIPQHLVEAEERVAEERVVM 180
DB      121 KDALKLLYSGVPLGPHKESAVTAFSEGSVIAYYSEFSIPQHLVEAEERVAEERVVM 180
QY      181 LPPRARSLSKSFVVTSSVAAPFTDSKTVOQTODNSCSFGLHARGVELMRFTTGPFPDSYPYA 240
DB      181 LPPRARSLSKSFVVTSSVAAPFTDSKTVOQTODNSCSFGLHARGVELMRFTTGPFPDSYPYA 240
QY      241 HARCOMALRGDADSVLSLTFRSFDLASCDERGSDLVTVYNTLSPMEPHALVOLCGTYPSPS 300
DB      241 HARCOMALRGDADSVLSLTFRSFDLASCDERGSDLVTVYNTLSPMEPHALVOLCGTYPSPS 300
QY      301 YNLTFFSSQNVLLITLITNTERRHGPGEATFFQLPRMSSCGGLRKAQGTFFNSPYYPGHY 360
DB      301 YNLTFFSSQNVLLITLITNTERRHGPGEATFFQLPRMSSCGGLRKAQGTFFNSPYYPGHY 360
QY      361 PPNIDCTWNIEVNNQHVKKRFFKFFYLLEPGVPAGTCPKDYVEINGEKYCGERSQFVYVS 420
DB      361 PPNIDCTWNIEVNNQHVKKRFFKFFYLLEPGVPAGTCPKDYVEINGEKYCGERSQFVYVS 420
QY      421 NSNKITVRFFSDOSYTDITGFLAEYLSYDSDPCPGQFTCTGTGCIKRELRCDMACTDH 480
DB      421 NSNKITVRFFSDOSYTDITGFLAEYLSYDSDPCPGQFTCTGTGCIKRELRCDMACTDH 480
QY      481 SDELINCSADAGHOFCTCKNFKCKPLFWVCDSVNDGSDSDEQSCSPAQTFRCSNGKCLSK 540
DB      481 SDELINCSADAGHOFCTCKNFKCKPLFWVCDSVNDGSDSDEQSCSPAQTFRCSNGKCLSK 540
QY      541 SQQCNMGDDCGSDSDEASCPKVVNVVTCTKTYRCLNGLCLSKNPPCEDGKEDCSGSDSK 600
DB      541 SQQCNMGDDCGSDSDEASCPKVVNVVTCTKTYRCLNGLCLSKNPPCEDGKEDCSGSDSK 600
QY      601 DCDGGLSFTROARVVGCTDADEGEWPMOYSLHALGQGHICGASLISPMNLVSAAHCIYD 660
```


Db 601 DDCGLRSFTROARVVGTDADBEEMFWQVSLHALGGCHICGASLISPMNLVSAHCYID 660
Qy 661 DRGFRYSDPQMTAFGLHDSQSRAPGVQERRLKRIISHPFNDFTFYDIALLELEKP 720
Db 661 DRGFRYSDPQMTAFGLHDSQSRAPGVQERRLKRIISHPFNDFTFYDIALLELEKP 720
Qy 721 AEYSMWPRICLPDASHVFPAGKAIWYTGWGHQYGGTGALILQKGEIRVINQTTCEML 780
Db 721 AEYSMWPRICLPDASHVFPAGKAIWYTGWGHQYGGTGALILQKGEIRVINQTTCEML 780
Qy 781 PQQITPRMVCVFLSGVDSCQDSGGLSVLEADGRIFGAGVSWGDGACGRNPGVYT 840
Db 781 PQQITPRMVCVFLSGVDSCQDSGGLSVLEADGRIFGAGVSWGDGACGRNPGVYT 840
Qy 841 RLPLFRDWIKENTGV 855
Db 841 RLPLFRDWIKENTGV 855

RESULT 13
US-10-112-221A-2
; Sequence 2, Application US/10112221A
; Publication No. US20030166851A1
; GENERAL INFORMATION:
; APPLICANT: Madison, Edwin
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING A TRANSMEMBRANE SERINE PROTEASE 9
; FILE REFERENCE: 24745-1615
; CURRENT APPLICATION NUMBER: US/10/112,221A
; PRIOR FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: 60/279,228
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: 60/291,501
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-112-221A-2

Query Match 99.4%; Score 4653; DB 14; Length 855;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 MGSBRARKGGGGPDPFAGLKYNSRHEKVNGLBEGVEFLPVNNVKKVEKHGPRWVTLAA 60
Db 1 MGSBRARKGGGGPDPFAGLKYNSRHEKVNGLBEGVEFLPVNNVKKVEKHGPRWVTLAA 60
Qy 61 VLIGLLVLTGIGFLVWHLQYRDVAVQKVKNGYMRITNENFVDAYENSNSTEFLVSLASKV 120
Db 61 VLIGLLVLTGIGFLVWHLQYRDVAVQKVKNGYMRITNENFVDAYENSNSTEFLVSLASKV 120
Qy 121 KDALKLTVSGVPLFGPYHKESAVTAFASEGSVIAYWSEFSIPQHLVEAEERVAEERVVM 180
Db 121 KDALKLTVSGVPLFGPYHKESAVTAFASEGSVIAYWSEFSIPQHLVEAEERVAEERVVM 180
Qy 181 LPPPARSLKSFVNVSVVAFPTDSKTVORTDONSCEFGLHAGVVELMRITTEGFPDSPVPA 240
Db 181 LPPPARSLKSFVNVSVVAFPTDSKTVORTDONSCEFGLHAGVVELMRITTEGFPDSPVPA 240
Qy 241 HARCOMALRGDADSVLTFRSFDLASCDERGSIDLVTYNTLSPMHPALVOLCGTYPSS 300
Db 241 HARCOMALRGDADSVLTFRSFDLASCDERGSIDLVTYNTLSPMHPALVOLCGTYPSS 300
Qy 301 YNLTFHSSQNVLLTLITNTERRHNGFEATFPQLPRMSSCGRLKAKAGCTRNSPYRPHY 360
Db 301 YNLTFHSSQNVLLTLITNTERRHNGFEATFPQLPRMSSCGRLKAKAGCTRNSPYRPHY 360
Qy 361 PPNIIDCNINIEVNNQHVKVPFKFYLLBPGVACTCPKDVYEINGEKYCGERSQFVVTSS 420

Db 361 PPNIIDCNINIEVNNQHVKVPFKFYLLBPGVACTCPKDVYEINGEKYCGERSQFVVTSS 420
Qy 421 NSNKITVRFHSDQSYTTTGFLAEVLSYDSDPCGQGTCTRGCIKRELRDGDMACTDTH 480
Db 421 NSNKITVRFHSDQSYTTTGFLAEVLSYDSDPCGQGTCTRGCIKRELRDGDMACTDTH 480
Qy 481 SDELNGSCDGHQTCCKNFKCKPLFWYCDVNDGSDNSDEQSCSPQOTFRCSNGKCLSK 540
Db 481 SDELNGSCDGHQTCCKNFKCKPLFWYCDVNDGSDNSDEQSCSPQOTFRCSNGKCLSK 540
Qy 541 SQQNGKDDCGDSDDEASCPVNVVCTKITYRCLNGLCISKNGPECDEKEDSDGSDK 600
Db 541 SQQNGKDDCGDSDDEASCPVNVVCTKITYRCLNGLCISKNGPECDEKEDSDGSDK 600
Qy 601 DDCGLRSFTROARVVGTDADBEEMFWQVSLHALGGCHICGASLISPMNLVSAHCYID 660
Db 601 DDCGLRSFTROARVVGTDADBEEMFWQVSLHALGGCHICGASLISPMNLVSAHCYID 660
Qy 661 DRGFRYSDPQMTAFGLHDSQSRAPGVQERRLKRIISHPFNDFTFYDIALLELEKP 720
Db 661 DRGFRYSDPQMTAFGLHDSQSRAPGVQERRLKRIISHPFNDFTFYDIALLELEKP 720
Qy 721 AEYSMWPRICLPDASHVFPAGKAIWYTGWGHQYGGTGALILQKGEIRVINQTTCEML 780
Db 721 AEYSMWPRICLPDASHVFPAGKAIWYTGWGHQYGGTGALILQKGEIRVINQTTCEML 780
Qy 781 PQQITPRMVCVFLSGVDSCQDSGGLSVLEADGRIFGAGVSWGDGACGRNPGVYT 840
Db 781 PQQITPRMVCVFLSGVDSCQDSGGLSVLEADGRIFGAGVSWGDGACGRNPGVYT 840
Qy 841 RLPLFRDWIKENTGV 855
Db 841 RLPLFRDWIKENTGV 855

RESULT 14
US-10-104-271-2
; Sequence 2, Application US/10104271
; Publication No. US20030181658A1
; GENERAL INFORMATION:
; APPLICANT: Madison, Edwin
; APPLICANT: Jiumn-Chern Yeh
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING SERINE PROTEASE CVSP14, THE ENCODI
; FILE REFERENCE: 24745-1614
; CURRENT APPLICATION NUMBER: US/10/104,271
; PRIOR FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: 60/278,166
; PRIOR FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-104-271-2

Query Match 99.4%; Score 4653; DB 14; Length 855;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 MGSBRARKGGGGPDPFAGLKYNSRHEKVNGLBEGVEFLPVNNVKKVEKHGPRWVTLAA 60
Db 1 MGSBRARKGGGGPDPFAGLKYNSRHEKVNGLBEGVEFLPVNNVKKVEKHGPRWVTLAA 60
Qy 61 VLIGLLVLTGIGFLVWHLQYRDVAVQKVKNGYMRITNENFVDAYENSNSTEFLVSLASKV 120
Db 61 VLIGLLVLTGIGFLVWHLQYRDVAVQKVKNGYMRITNENFVDAYENSNSTEFLVSLASKV 120
Qy 121 KDALKLTVSGVPLFGPYHKESAVTAFASEGSVIAYWSEFSIPQHLVEAEERVAEERVVM 180
Db 121 KDALKLTVSGVPLFGPYHKESAVTAFASEGSVIAYWSEFSIPQHLVEAEERVAEERVVM 180

```
QY 181 LPPRSLKSFVVTSVVAFPTDSKTQRTODNSCSFGLHARGVELMFTTGPDPSPYA 240
|
|
|
Db 181 LPPRSLKSFVVTSVVAFPTDSKTQRTODNSCSFGLHARGVELMFTTGPDPSPYA 240
QY 241 HARCQMLRGDADSVLSLTFRSFDLASCDERGSPLVTVYNTLSPMEPHALVOLCGTYP 300
|
|
|
Db 241 HARCQMLRGDADSVLSLTFRSFDLASCDERGSPLVTVYNTLSPMEPHALVOLCGTYP 300
QY 301 YNLTFHSSQVLLITLITNTERHHPGEATFFQLPRMSSCGRLRKAQGTNSPYPGHY 360
|
|
|
Db 301 YNLTFHSSQVLLITLITNTERHHPGEATFFQLPRMSSCGRLRKAQGTNSPYPGHY 360
QY 361 PPNIDCTWNIIEVNNQHVKRFKFFYLLEPGVAGTCKPDYVEINERKXCGERSQFVYTS 420
|
|
|
Db 361 PPNIDCTWNIIEVNNQHVKRFKFFYLLEPGVAGTCKPDYVEINERKXCGERSQFVYTS 420
QY 421 NSNKITVRFHSDQSYDTGTGLAEYLSYSSDPGQFTCRGRCIRKELRCDGMADCTDH 480
|
|
|
Db 421 NSNKITVRFHSDQSYDTGTGLAEYLSYSSDPGQFTCRGRCIRKELRCDGMADCTDH 480
QY 481 SDELNCSGADGHOFTCKNFKCKPLFWYCDSVNDCGNSDEQSCSPAQTFRCNSGKCLSK 540
|
|
|
Db 481 SDELNCSGADGHOFTCKNFKCKPLFWYCDSVNDCGNSDEQSCSPAQTFRCNSGKCLSK 540
QY 541 SQQCNKGKDCGSDSDEASCPKRVNVVCTKHTYRCLNGLCLSKNPECDGKEDCSGSD 600
|
|
|
Db 541 SQQCNKGKDCGSDSDEASCPKRVNVVCTKHTYRCLNGLCLSKNPECDGKEDCSGSD 600
QY 601 DCCGGLRSTFRQARVVGCTDADEGEWPWQVSLHALCGGHI CGASLISPMVLVSAAHCT 660
|
|
|
Db 601 DCCGGLRSTFRQARVVGCTDADEGEWPWQVSLHALCGGHI CGASLISPMVLVSAAHCT 660
QY 661 DRGRYSDDPTQMTAFGLHDSQSRAPGVQERLKRILISHPFNDFFDYDIALLELEK 720
|
|
|
Db 661 DRGRYSDDPTQMTAFGLHDSQSRAPGVQERLKRILISHPFNDFFDYDIALLELEK 720
QY 721 AEYSMWPRICLPDASHVFPAGKAIWVTGWHGTQYGGTGALILQKGEIRVINQTTCE 780
|
|
|
Db 721 AEYSMWPRICLPDASHVFPAGKAIWVTGWHGTQYGGTGALILQKGEIRVINQTTCE 780
QY 781 PQQITPRMVCVGLSGVDSQCGSGPLSSVEADGRIFGAGVSWGDCAGRNKGVYT 840
|
|
|
Db 781 PQQITPRMVCVGLSGVDSQCGSGPLSSVEADGRIFGAGVSWGDCAGRNKGVYT 840
QY 841 RLPLFRDMIKENTGV 855
|
|
|
Db 841 RLPLFRDMIKENTGV 855

RESULT 15
US-10-147-211A-2
; Sequence 2, Application US/10147211A
; Publication No. US20030235900A1
; GENERAL INFORMATION:
; APPLICANT: Madison, Edward
; APPLICANT: Yeh, Jiumn-Chern
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING A TRANSMEMBRANE SERINE PROTEASE 1
; FILE REFERENCE: 24745-1616
; CURRENT APPLICATION NUMBER: US/10/147,211A
; CURRENT FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/291,001
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-147-211A-2
Query Match 99.4%; Score 4653; DB 14; Length 855;
```

```
Best Local Similarity 99.5%; Pred. No. 0;
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 MSDBRARKGGGKPDGAGLKYNRSREKXVNGLEGEVLEPVNNVKKVEKHPRWVLLAA 60
|
|
|
Db 1 MSDBRARKGGGKPDGAGLKYNRSREKXVNGLEGEVLEPVNNVKKVEKHPRWVLLAA 60
QY 61 VILGLLVLLIGIFLVNHLQYRDVRYQKXNNGYRITNENFVADYENSSTEFVSLASKV 120
|
|
|
Db 61 VILGLLVLLIGIFLVNHLQYRDVRYQKXNNGYRITNENFVADYENSSTEFVSLASKV 120
QY 121 KDLKLLYSGVPLGYPHKESAVTAASEGVIAIYVNSEPSIPQHLVEEARVMAEERVVM 180
|
|
|
Db 121 KDLKLLYSGVPLGYPHKESAVTAASEGVIAIYVNSEPSIPQHLVEEARVMAEERVVM 180
QY 181 LPPRSLKSFVVTSVVAFPTDSKTQRTODNSCSFGLHARGVELMFTTGPDPSPYA 240
|
|
|
Db 181 LPPRSLKSFVVTSVVAFPTDSKTQRTODNSCSFGLHARGVELMFTTGPDPSPYA 240
QY 241 HARCQMLRGDADSVLSLTFRSFDLASCDERGSPLVTVYNTLSPMEPHALVOLCGTYP 300
|
|
|
Db 241 HARCQMLRGDADSVLSLTFRSFDLASCDERGSPLVTVYNTLSPMEPHALVOLCGTYP 300
QY 301 YNLTFHSSQVLLITLITNTERHHPGEATFFQLPRMSSCGRLRKAQGTNSPYPGHY 360
|
|
|
Db 301 YNLTFHSSQVLLITLITNTERHHPGEATFFQLPRMSSCGRLRKAQGTNSPYPGHY 360
QY 361 PPNIDCTWNIIEVNNQHVKRFKFFYLLEPGVAGTCKPDYVEINERKXCGERSQFVYTS 420
|
|
|
Db 361 PPNIDCTWNIIEVNNQHVKRFKFFYLLEPGVAGTCKPDYVEINERKXCGERSQFVYTS 420
QY 421 NSNKITVRFHSDQSYDTGTGLAEYLSYSSDPGQFTCRGRCIRKELRCDGMADCTDH 480
|
|
|
Db 421 NSNKITVRFHSDQSYDTGTGLAEYLSYSSDPGQFTCRGRCIRKELRCDGMADCTDH 480
QY 481 SDELNCSGADGHOFTCKNFKCKPLFWYCDSVNDCGNSDEQSCSPAQTFRCNSGKCLSK 540
|
|
|
Db 481 SDELNCSGADGHOFTCKNFKCKPLFWYCDSVNDCGNSDEQSCSPAQTFRCNSGKCLSK 540
QY 541 SQQCNKGKDCGSDSDEASCPKRVNVVCTKHTYRCLNGLCLSKNPECDGKEDCSGSD 600
|
|
|
Db 541 SQQCNKGKDCGSDSDEASCPKRVNVVCTKHTYRCLNGLCLSKNPECDGKEDCSGSD 600
QY 601 DCCGGLRSTFRQARVVGCTDADEGEWPWQVSLHALCGGHI CGASLISPMVLVSAAHCT 660
|
|
|
Db 601 DCCGGLRSTFRQARVVGCTDADEGEWPWQVSLHALCGGHI CGASLISPMVLVSAAHCT 660
QY 661 DRGRYSDDPTQMTAFGLHDSQSRAPGVQERLKRILISHPFNDFFDYDIALLELEK 720
|
|
|
Db 661 DRGRYSDDPTQMTAFGLHDSQSRAPGVQERLKRILISHPFNDFFDYDIALLELEK 720
QY 721 AEYSMWPRICLPDASHVFPAGKAIWVTGWHGTQYGGTGALILQKGEIRVINQTTCE 780
|
|
|
Db 721 AEYSMWPRICLPDASHVFPAGKAIWVTGWHGTQYGGTGALILQKGEIRVINQTTCE 780
QY 781 PQQITPRMVCVGLSGVDSQCGSGPLSSVEADGRIFGAGVSWGDCAGRNKGVYT 840
|
|
|
Db 781 PQQITPRMVCVGLSGVDSQCGSGPLSSVEADGRIFGAGVSWGDCAGRNKGVYT 840
QY 841 RLPLFRDMIKENTGV 855
|
|
|
Db 841 RLPLFRDMIKENTGV 855

RESULT 16
US-10-156-214A-2
; Sequence 2, Application US/10156214A
; Publication No. US20040001801A1
; GENERAL INFORMATION:
; APPLICANT: Edwin L. Madison
; APPLICANT: Joseph Edward Semple
; APPLICANT: George P. Vlaauk
; APPLICANT: Scott Jeffrey Kemp
```

APPLICANT: Mallareddy Komandla
APPLICANT: Daniel Vanna Siev
TITLE OF INVENTION: Conjugates Activated By Cell Surface Proteases and Therapeutic Use
TITLE OF INVENTION: Theeotf
FILE REFERENCE: 24745-1611
CURRENT APPLICATION NUMBER: US/10/156,214A
CURRENT FILING DATE: 2002-05-23
NUMBER OF SEQ ID NOS: 611
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 855
TYPE: PRT
ORGANISM: Homo Sapien
US-10-156-214A-2

Query Match 99.4%; Score 4653; DB 15; Length 855;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

1 MGSDRAKGGGKDPFGAGLKYNSRHEKXNGLEGEVEFLPVNNVKKYKKGPGRWVTLAA 60
1 MGSDRAKGGGKDPFGAGLKYNSRHEKXNGLEGEVEFLPVNNVKKYKKGPGRWVTLAA 60
61 VLIGLLVLLIGLFLVWHLQYRDYRVOVKXKXGYNRTINENFVADYENSNGTEFVSLASKV 120
61 VLIGLLVLLIGLFLVWHLQYRDYRVOVKXKXGYNRTINENFVADYENSNGTEFVSLASKV 120
121 KDALKLISGVPLGPRHKSAAVAFBSGVIATYWSBPSIPQHLVEBAEYVMAEEYVM 180
121 KDALKLISGVPLGPRHKSAAVAFBSGVIATYWSBPSIPQHLVEBAEYVMAEEYVM 180
121 KDALKLISGVPLGPRHKSAAVAFBSGVIATYWSBPSIPQHLVEBAEYVMAEEYVM 180
181 LPPRARSLSFVWTSVVAAPFDSKTQVORTDONSFGIHAAGVLMFTTGFPSDPYPA 240
181 LPPRARSLSFVWTSVVAAPFDSKTQVORTDONSFGIHAAGVLMFTTGFPSDPYPA 240
181 LPPRARSLSFVWTSVVAAPFDSKTQVORTDONSFGIHAAGVLMFTTGFPSDPYPA 240
241 HARQWALRGDADSVLSLTFRSPDLASCDERGSDLVTYNTLSMEPHALVOLGTYPPS 300
241 HARQWALRGDADSVLSLTFRSPDLASCDERGSDLVTYNTLSMEPHALVOLGTYPPS 300
301 YNLTFRSSQVNLTLITNTERRRPGEATFPQIPRNSCCGGRKAKQGTNSYYGXY 360
301 YNLTFRSSQVNLTLITNTERRRPGEATFPQIPRNSCCGGRKAKQGTNSYYGXY 360
301 YNLTFRSSQVNLTLITNTERRRPGEATFPQIPRNSCCGGRKAKQGTNSYYGXY 360
361 PPNIDCWNTLVNPNQHVKKRFFKFFYLBPVPAGTCKDVEINGEYCGERSQFVYTS 420
361 PPNIDCWNTLVNPNQHVKKRFFKFFYLBPVPAGTCKDVEINGEYCGERSQFVYTS 420
361 PPNIDCWNTLVNPNQHVKKRFFKFFYLBPVPAGTCKDVEINGEYCGERSQFVYTS 420
421 NSNKTIVRFHSDSYTDGFLAEYLSYDSDPCGQFTCTGRCIRKELRCGMACTDH 480
421 NSNKTIVRFHSDSYTDGFLAEYLSYDSDPCGQFTCTGRCIRKELRCGMACTDH 480
421 NSNKTIVRFHSDSYTDGFLAEYLSYDSDPCGQFTCTGRCIRKELRCGMACTDH 480
481 SDELINSCDAGHQTCKKCKCPFLWVCDVNDGSDNSDEGSCCPQOTFRCSNGKLSK 540
481 SDELINSCDAGHQTCKKCKCPFLWVCDVNDGSDNSDEGSCCPQOTFRCSNGKLSK 540
481 SDELINSCDAGHQTCKKCKCPFLWVCDVNDGSDNSDEGSCCPQOTFRCSNGKLSK 540
541 SQQNGSDGDDGSDDEASCPKVVNVVCTKHTYRCLNGLCLSKGPECDGKSDGSDG 600
541 SQQNGSDGDDGSDDEASCPKVVNVVCTKHTYRCLNGLCLSKGPECDGKSDGSDG 600
541 SQQNGSDGDDGSDDEASCPKVVNVVCTKHTYRCLNGLCLSKGPECDGKSDGSDG 600
601 DCDGGLSFTROARVVGTDADGEMPMOYSLHALGCGHICGASLISPMNLVSAHICYD 660
601 DCDGGLSFTROARVVGTDADGEMPMOYSLHALGCGHICGASLISPMNLVSAHICYD 660
601 DCDGGLSFTROARVVGTDADGEMPMOYSLHALGCGHICGASLISPMNLVSAHICYD 660
661 DRGRYSDDPTQWTAFLGLHDSQSRAPGVQERRLRIISHPFNDFTFYDIALLEKXP 720
661 DRGRYSDDPTQWTAFLGLHDSQSRAPGVQERRLRIISHPFNDFTFYDIALLEKXP 720
661 DRGRYSDDPTQWTAFLGLHDSQSRAPGVQERRLRIISHPFNDFTFYDIALLEKXP 720
721 AYSWVRPCLPASHVPPAKAIWYTGWHTOYGGALILQKGRIVNNTTCENL 780
721 AYSWVRPCLPASHVPPAKAIWYTGWHTOYGGALILQKGRIVNNTTCENL 780
721 AYSWVRPCLPASHVPPAKAIWYTGWHTOYGGALILQKGRIVNNTTCENL 780
781 PQQITPRMVCVFLSGVDSCQDGGPLSSVEADGRIFGAGVVSWDGCAQRNKPQVYT 840
781 PQQITPRMVCVFLSGVDSCQDGGPLSSVEADGRIFGAGVVSWDGCAQRNKPQVYT 840
781 PQQITPRMVCVFLSGVDSCQDGGPLSSVEADGRIFGAGVVSWDGCAQRNKPQVYT 840

Db 781 PQQITPRMVCVFLSGVDSCQDGGPLSSVEADGRIFGAGVVSWDGCAQRNKPQVYT 840
Qy 841 RLPLFRMIXENTGV 855
Db 841 RLPLFRMIXENTGV 855

RESULT 17
US-10-072-012-352

Sequence 352, Application US/10072012
Publication No. US2004003493A1
GENERAL INFORMATION:
APPLICANT: Tchernyev, Velizar
APPLICANT: Spytek, Kimberly
APPLICANT: Zettersen, Bryan
APPLICANT: Paturajan, Meera
APPLICANT: Shukets, Richard
APPLICANT: Li, Li
APPLICANT: Gangoli, Beha
APPLICANT: Padigara, Muralidhara
APPLICANT: Anderson, David W.
APPLICANT: Rastelli, Luca
APPLICANT: Miller, Charles E.
APPLICANT: Gerlach, Valerie
APPLICANT: Taupier Jr, Raymond J.
APPLICANT: Gusev, Vladimir Y.
APPLICANT: Coleman, Steven D.
APPLICANT: Wolenc, Adam R.
APPLICANT: Pena, Carol E. A
APPLICANT: Furtak, Katarzyna
APPLICANT: Grose, William M.
APPLICANT: Alebrock II, John P.
APPLICANT: Lepley, Denise M.
APPLICANT: Rieger, Daniel K.
APPLICANT: Burgess, Catherine E.
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
FILE REFERENCE: 21402-258
CURRENT APPLICATION NUMBER: US/10/072,012
PRIOR FILING DATE: 2002-01-31
PRIOR APPLICATION NUMBER: 60/265,102
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: 60/265,514
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265,517
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265,412
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265,395
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/266,406
PRIOR FILING DATE: 2001-02-02
PRIOR APPLICATION NUMBER: 60/266,767
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: 60/267,057
PRIOR FILING DATE: 2001-02-07
PRIOR APPLICATION NUMBER: 60/266,975
PRIOR FILING DATE: 2001-02-07
PRIOR APPLICATION NUMBER: 60/267,459
PRIOR FILING DATE: 2001-02-08
Remainder Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 1391
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 352
LENGTH: 855
TYPE: PRT
ORGANISM: Homo sapiens
US-10-072-012-352

Query Match 99.4%; Score 4653; DB 15; Length 855;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
1 MGSDRAKGGGKDPFGAGLKYNSRHEKXNGLEGEVEFLPVNNVKKYKKGPGRWVTLAA 60

```

Db      1  MGSDRARKGGGKPDGAGLKYNSRHEKYNGLBEGVEFLPVNNVKYKKGPRWVLLAA 60
Qy      61  VLLGLLVLLIGIFLVWHLQYRDVRYQKYNQYWRINENFVAAYENSNEFEVSLASKY 120
Db      61  VLLGLLVLLIGIFLVWHLQYRDVRYQKYNQYWRINENFVAAYENSNEFEVSLASKY 120
Qy      121  KDALKLLYSGVPLGPGYHKEASAVTAFSEGSVIAAYMSEFSI PQHLYEAEERVAEERVM 180
Db      121  KDALKLLYSGVPLGPGYHKEASAVTAFSEGSVIAAYMSEFSI PQHLYEAEERVAEERVM 180
Qy      181  LPPRARSLSFVVTSVVAFPTDSKTIVRTQDNGSCFGLHARGVELMRFTTPGPDSPYPA 240
Db      181  LPPRARSLSFVVTSVVAFPTDSKTIVRTQDNGSCFGLHARGVELMRFTTPGPDSPYPA 240
Qy      241  HARCQWALRGDADSVSLTFRSPDLASCDERGSDLVTYVNTLSPMEPHALVOLCGTYPSP 300
Db      241  HARCQWALRGDADSVSLTFRSPDLASCDERGSDLVTYVNTLSPMEPHALVOLCGTYPSP 300
Qy      301  YNLTFHSSQNVLLITLITNTERRHPGFEATFPOLPRMSSCGRLRKAQGTENSPYYPGHY 360
Db      301  YNLTFHSSQNVLLITLITNTERRHPGFEATFPOLPRMSSCGRLRKAQGTENSPYYPGHY 360
Qy      361  PPNIIDCTWNIENVNNOYKVRKPKFYLLBEGVPACTPKQYVEINRGKTCGERSQFYVTS 420
Db      361  PPNIIDCTWNIENVNNOYKVRKPKFYLLBEGVPACTPKQYVEINRGKTCGERSQFYVTS 420
Qy      421  NSNKITVRFHSDOSYDTGTGFLAEIAYSDDSPCGQGTCTGRGRCIRRELRCDDGADCTDH 480
Db      421  NSNKITVRFHSDOSYDTGTGFLAEIAYSDDSPCGQGTCTGRGRCIRRELRCDDGADCTDH 480
Qy      481  SDEINSCDAGHOFCTCKNKKCPPLFWVCDSDVNDGDNDSDEQSCCPAQTRFCSNGKCLSK 540
Db      481  SDEINSCDAGHOFCTCKNKKCPPLFWVCDSDVNDGDNDSDEQSCCPAQTRFCSNGKCLSK 540
Qy      541  SQQNGDDDCGDSDEASCPKRVNVTCTKTYRCLNGLCLSKNPEDEGKEDSDGSDXK 600
Db      541  SQQNGDDDCGDSDEASCPKRVNVTCTKTYRCLNGLCLSKNPEDEGKEDSDGSDXK 600
Qy      601  DCCGGLSFTROARVVGTDADBEEMPQVSLHLAGGHI CGASLISPNMLVSAHACYID 660
Db      601  DCCGGLSFTROARVVGTDADBEEMPQVSLHLAGGHI CGASLISPNMLVSAHACYID 660
Qy      661  DRGFRYSPTQWTAFLGLHDQSORAPGVQERRLKRIISHPFNDFTFDYDIALLELEKP 720
Db      661  DRGFRYSPTQWTAFLGLHDQSORAPGVQERRLKRIISHPFNDFTFDYDIALLELEKP 720
Qy      721  AEYSWWRPCLDASIVFPAGKAIWTTGHTQYGGTGALLIQKGIIRYNQTCENL 780
Db      721  AEYSWWRPCLDASIVFPAGKAIWTTGHTQYGGTGALLIQKGIIRYNQTCENL 780
Qy      781  PQGITPRMNCVGLSGVDSCQSGPLSSVEADGRI FGAQVSWMGDGCAGRNKPPVYT 840
Db      781  PQGITPRMNCVGLSGVDSCQSGPLSSVEADGRI FGAQVSWMGDGCAGRNKPPVYT 840
Qy      841  RLPLFRDMIKENTGV 855
Db      841  RLPLFRDMIKENTGV 855

```

```

; APPLICANT: Anderson, David W.
; APPLICANT: Rastelli, Luca
; APPLICANT: Miller, Charles E.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Molenc, Steven D.
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Grose, William M.
; APPLICANT: Alsobrook II, John P.
; APPLICANT: Lepley, Denise M.
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Burgess, Catherine E.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-258
; CURRENT APPLICATION NUMBER: US/10/072, 012
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 60/265,102
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/265,514
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,412
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,395
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/266,406
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/266,767
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/267,057
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/266,975
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/267,459
; PRIOR FILING DATE: 2001-02-08
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1391
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 411
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-072-012-411

Query Match      99.4%; Score 4653; DB 15; Length 855;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy      1  MGSDRARKGGGKPDGAGLKYNSRHEKYNGLBEGVEFLPVNNVKYKKGPRWVLLAA 60
Db      1  MGSDRARKGGGKPDGAGLKYNSRHEKYNGLBEGVEFLPVNNVKYKKGPRWVLLAA 60
Qy      61  VLLGLLVLLIGIFLVWHLQYRDVRYQKYNQYWRINENFVAAYENSNEFEVSLASKY 120
Db      61  VLLGLLVLLIGIFLVWHLQYRDVRYQKYNQYWRINENFVAAYENSNEFEVSLASKY 120
Qy      121  KDALKLLYSGVPLGPGYHKEASAVTAFSEGSVIAAYMSEFSI PQHLYEAEERVAEERVM 180
Db      121  KDALKLLYSGVPLGPGYHKEASAVTAFSEGSVIAAYMSEFSI PQHLYEAEERVAEERVM 180
Qy      181  LPPRARSLSFVVTSVVAFPTDSKTIVRTQDNGSCFGLHARGVELMRFTTPGPDSPYPA 240
Db      181  LPPRARSLSFVVTSVVAFPTDSKTIVRTQDNGSCFGLHARGVELMRFTTPGPDSPYPA 240
Qy      241  HARCQWALRGDADSVSLTFRSPDLASCDERGSDLVTYVNTLSPMEPHALVOLCGTYPSP 300
Db      241  HARCQWALRGDADSVSLTFRSPDLASCDERGSDLVTYVNTLSPMEPHALVOLCGTYPSP 300
Qy      301  YNLTFHSSQNVLLITLITNTERRHPGFEATFPOLPRMSSCGRLRKAQGTENSPYYPGHY 360

```

```
DB 301 YNLFHSSQVNLITLITNTERRHPGEATFPOLPRMSSCGRLRKAQGTFFNSYYTGHY 360
QY 361 PENIDCTWNIIEVNNQHVKVRKFKFYLLBPGVAGTCKPDYVEINGEKYCGERSQFVVT 420
DB 361 PENIDCTWNIIEVNNQHVKVSFKFYLLBPGVAGTCKPDYVEINGEKYCGERSQFVVT 420
QY 421 NSNKITRFRHSDDSYTDTGFLAEYLSYDSDPCPGQFTCTGRCIRKELACDGMADCTDH 480
DB 421 NSNKITRFRHSDDSYTDTGFLAEYLSYDSDPCPGQFTCTGRCIRKELACDGMADCTDH 480
QY 481 SDELINSCDAGHOFCTCKNKECKPLFWYCDVNDGSDNDEOGSCCPAQTRFRCNGKCLSK 540
DB 481 SDELINSCDAGHOFCTCKNKECKPLFWYCDVNDGSDNDEOGSCCPAQTRFRCNGKCLSK 540
QY 541 SOQCNKGKDDCGSDSDEASCPKRVNVVTCTKATYRCLNGLCLSKNPECDCGKEDSDGSEK 600
DB 541 SOQCNKGKDDCGSDSDEASCPKRVNVVTCTKATYRCLNGLCLSKNPECDCGKEDSDGSEK 600
QY 601 DCCGGLRSFTRQARVVGTDADBEEMWQVSLHALGQGHICGASLISPNMLVSAHACYID 660
DB 601 DCCGGLRSFTRQARVVGTDADBEEMWQVSLHALGQGHICGASLISPNMLVSAHACYID 660
QY 661 DRGFRYSDPQMTAFGLGHDOSRARPVOERRLKRIISHPFNDPFEDYDIALLELEKXP 720
DB 661 DRGFRYSDPQMTAFGLGHDOSRARPVOERRLKRIISHPFNDPFEDYDIALLELEKXP 720
QY 721 AEYSMWVRP1CLPDASHVPFAGKAIWVTGHTQYGGTGALLQKGEIRVYNQTTCEML 780
DB 721 AEYSMWVRP1CLPDASHVPFAGKAIWVTGHTQYGGTGALLQKGEIRVYNQTTCEML 780
QY 781 POQITPRMVCVGLSGVDSCQDSCGRLSVEADGILFGAGVYVWMDGCAQRKRPVYT 840
DB 781 POQITPRMVCVGLSGVDSCQDSCGRLSVEADGILFGAGVYVWMDGCAQRKRPVYT 840
QY 841 RLPLFRDMIKENTGV 855
DB 841 RLPLFRDMIKENTGV 855

RESULT 19
; Sequence 418, Application US/10072012
; Publication No. US2004003493A1
; GENERAL INFORMATION:
; APPLICANT: Tchernyev, Velizar
; APPLICANT: Spyttek, Kimberly
; APPLICANT: Zetshusen, Bryan
; APPLICANT: Patcurajan, Meera
; APPLICANT: Shumkets, Richard
; APPLICANT: Li, Li
; APPLICANT: Gangolli, Esha
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Anderson, David W.
; APPLICANT: Rastelli, Luca
; APPLICANT: Miller, Charles E.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Colman, Steven D.
; APPLICANT: Wolenc, Adam R.
; APPLICANT: Pena, Carol E. A
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Grose, William M.
; APPLICANT: Aleobrook II, John P.
; APPLICANT: Lepley, Denise M.
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Burgess, Catherine E.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-258
; CURRENT APPLICATION NUMBER: US/10/072.012
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 60/265,102
```

```
QY 1 MSDDRARKGGGKPKDAGLCKYNSRHEKNGLEEGYEPPLPVNNVYKVEKGPGRWVLLA 60
DB 1 MSDDRARKGGGKPKDAGLCKYNSRHEKNGLEEGYEPPLPVNNVYKVEKGPGRWVLLA 60
QY 61 VILGLLVILGLIGFLVWHLOYRDVRYOKYNGYMRITNENFVDAYENSNTSEFVSLASKV 120
DB 61 VILGLLVILGLIGFLVWHLOYRDVRYOKYNGYMRITNENFVDAYENSNTSEFVSLASKV 120
QY 121 KDALKLLYSGVPLGPGYHKSATVAFSEGSVIAVYSEFISIPQHLVEAERVAEERVVM 180
DB 121 KDALKLLYSGVPLGPGYHKSATVAFSEGSVIAVYSEFISIPQHLVEAERVAEERVVM 180
QY 181 LPPRASLKSFTVTSVVAFTDSKTYQRTODNCSGFLHARGVELMRFTTPGPPDPYPA 240
DB 181 LPPRASLKSFTVTSVVAFTDSKTYQRTODNCSGFLHARGVELMRFTTPGPPDPYPA 240
QY 241 HARCQWALRGDADSVLSLFRSFDLASCDERGSDLTVVYNTLSPMEPHALVOLCGTYPPS 300
DB 241 HARCQWALRGDADSVLSLFRSFDLASCDERGSDLTVVYNTLSPMEPHALVOLCGTYPPS 300
QY 301 YNLFHSSQVNLITLITNTERRHPGEATFPOLPRMSSCGRLRKAQGTFFNSYYTGHY 360
DB 301 YNLFHSSQVNLITLITNTERRHPGEATFPOLPRMSSCGRLRKAQGTFFNSYYTGHY 360
QY 361 PENIDCTWNIIEVNNQHVKVRKFKFYLLBPGVAGTCKPDYVEINGEKYCGERSQFVVT 420
DB 361 PENIDCTWNIIEVNNQHVKVSFKFYLLBPGVAGTCKPDYVEINGEKYCGERSQFVVT 420
QY 421 NSNKITRFRHSDDSYTDTGFLAEYLSYDSDPCPGQFTCTGRCIRKELACDGMADCTDH 480
DB 421 NSNKITRFRHSDDSYTDTGFLAEYLSYDSDPCPGQFTCTGRCIRKELACDGMADCTDH 480
QY 481 SDELINSCDAGHOFCTCKNKECKPLFWYCDVNDGSDNDEOGSCCPAQTRFRCNGKCLSK 540
DB 481 SDELINSCDAGHOFCTCKNKECKPLFWYCDVNDGSDNDEOGSCCPAQTRFRCNGKCLSK 540
QY 541 SOQCNKGKDDCGSDSDEASCPKRVNVVTCTKATYRCLNGLCLSKNPECDCGKEDSDGSEK 600
DB 541 SOQCNKGKDDCGSDSDEASCPKRVNVVTCTKATYRCLNGLCLSKNPECDCGKEDSDGSEK 600
QY 601 DCCGGLRSFTRQARVVGTDADBEEMWQVSLHALGQGHICGASLISPNMLVSAHACYID 660
```

Db 601 DCCGRLSFTQARVVGTTADDEGEMPMQVSLHALGQGHICGASLISPMWLVSAAHCYID 660
Qy 661 DRGFRYSDFPTQWTAFLGLHDQSORSAPOVERLRKRIISHPFNDFTFYDIALLELEXP 720
Db 661 DRGFRYSDFPTQWTAFLGLHDQSORSAPOVERLRKRIISHPFNDFTFYDIALLELEXP 720
Qy 721 AEYSMWVRPCLPDASHVPAGKAIWVTGMGHTQYGGTGALLIOKGEIRYINOTTCENLL 780
Db 721 AEYSMWVRPCLPDASHVPAGKAIWVTGMGHTQYGGTGALLIOKGEIRYINOTTCENLL 780
Qy 781 PQGITPRMNCVGLSGGVDSQQDSSGGLSSVEADGRIFGAGVSWGDCAGRNKPGVYT 840
Db 781 PQGITPRMNCVGLSGGVDSQQDSSGGLSSVEADGRIFGAGVSWGDCAGRNKPGVYT 840
Qy 841 RLPLFRDWIKENTGV 855
Db 841 RLPLFRDWIKENTGV 855

RESULT 20

US-10-600-187-2
; Sequence 2, Application US/10600187
; Publication No. US20040086910A1
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Tanimoto, Hirotsoshi
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease
; TITLE OF INVENTION: Overexpressed in Carcinomas
; FILE REFERENCE: D6064CIF/D
; CURRENT APPLICATION NUMBER: US/10/600,187
; CURRENT FILING DATE: 2003-06-20
; PRIOR APPLICATION NUMBER: US/09/654,600A
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 09/421,213
; 09/027,337
; PRIOR FILING DATE: 1999-10-20
; 1998-02-20
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: TADG-15
US-10-600-187-2

Query Match 99.4%; Score 4653; DB 15; Length 855;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 MGSDBARKGGGGKDFGAGIKYNSRHEKVNGLGEVFLPVNNVKKVEKKGPRWVTLAA 60
Db 1 MGSDBARKGGGGKDFGAGIKYNSRHEKVNGLGEVFLPVNNVKKVEKKGPRWVTLAA 60
Qy 61 VLLGLLVLLIGIFLVWHLQYRDVRYQKXNGYMRITNENFVDAYENSNSTEFVSLASKV 120
Db 61 VLLGLLVLLIGIFLVWHLQYRDVRYQKXNGYMRITNENFVDAYENSNSTEFVSLASKV 120
Qy 121 KDLAKLLYSVPRFLGPHKKSANTAFSEGSVIAYMSEFSIPHLVEAEERVAEERVVM 180
Db 121 KDLAKLLYSVPRFLGPHKKSANTAFSEGSVIAYMSEFSIPHLVEAEERVAEERVVM 180
Qy 181 LPPRASLSKSFVYTSVVAFPDTSKTVQRTQDNCSFGLHARGVLEWFTTGPFPDSPYPA 240
Db 181 LPPRASLSKSFVYTSVVAFPDTSKTVQRTQDNCSFGLHARGVLEWFTTGPFPDSPYPA 240
Qy 241 HARCQWALRGDADSVSLTFERSFDLASCDBRGSDLVVYNTLSPMEPHALVOLGTYPPS 300
Db 241 HARCQWALRGDADSVSLTFERSFDLASCDBRGSDLVVYNTLSPMEPHALVOLGTYPPS 300
Qy 301 YNLTFFHSQVALLITLITNTERRHGFEATFPQLPRMSSCGRLRKAQGTFSNPPYPGHY 360
Db 301 YNLTFFHSQVALLITLITNTERRHGFEATFPQLPRMSSCGRLRKAQGTFSNPPYPGHY 360

Db 301 YNLTFFHSQVALLITLITNTERRHGFEATFPQLPRMSSCGRLRKAQGTFSNPPYPGHY 360
Qy 361 PPNIDCTMNIIEVNNOHVVRPFKFFYLLEPGVAGTCCPDYVEINGEKTCGERSQFVYTS 420
Db 361 PPNIDCTMNIIEVNNOHVVRPFKFFYLLEPGVAGTCCPDYVEINGEKTCGERSQFVYTS 420
Qy 421 NSNKITVRPHSDQSYTDTGFLAEYLSYDSSDPGQFTCRGTGCIKRELRCGDMADCTDH 480
Db 421 NSNKITVRPHSDQSYTDTGFLAEYLSYDSSDPGQFTCRGTGCIKRELRCGDMADCTDH 480
Qy 481 SDELNCSDAHQDFTCKNFKCPLFWVCDSVNDCCGNSDEQSCSCPAQTFRCNGKCLSK 540
Db 481 SDELNCSDAHQDFTCKNFKCPLFWVCDSVNDCCGNSDEQSCSCPAQTFRCNGKCLSK 540
Qy 541 SQQCNKDCDQSDSDASCPKXNVVTCYKHTYRCLNGLCLSKGNPCDQKEDCSDSDSK 600
Db 541 SQQCNKDCDQSDSDASCPKXNVVTCYKHTYRCLNGLCLSKGNPCDQKEDCSDSDSK 600
Qy 601 DCDGRLSFTQARVVGTTADDEGEMPMQVSLHALGQGHICGASLISPMWLVSAAHCYID 660
Db 601 DCDGRLSFTQARVVGTTADDEGEMPMQVSLHALGQGHICGASLISPMWLVSAAHCYID 660
Qy 661 DRGFRYSDFPTQWTAFLGLHDQSORSAPOVERLRKRIISHPFNDFTFYDIALLELEXP 720
Db 661 DRGFRYSDFPTQWTAFLGLHDQSORSAPOVERLRKRIISHPFNDFTFYDIALLELEXP 720
Qy 721 AEYSMWVRPCLPDASHVPAGKAIWVTGMGHTQYGGTGALLIOKGEIRYINOTTCENLL 780
Db 721 AEYSMWVRPCLPDASHVPAGKAIWVTGMGHTQYGGTGALLIOKGEIRYINOTTCENLL 780
Qy 781 PQGITPRMNCVGLSGGVDSQQDSSGGLSSVEADGRIFGAGVSWGDCAGRNKPGVYT 840
Db 781 PQGITPRMNCVGLSGGVDSQQDSSGGLSSVEADGRIFGAGVSWGDCAGRNKPGVYT 840
Qy 841 RLPLFRDWIKENTGV 855
Db 841 RLPLFRDWIKENTGV 855

RESULT 21

US-10-276-774-1798
; Sequence 1798, Application US/10276774
; Publication No. US20040053245A1
; GENERAL INFORMATION:
; APPLICANT: Hyeeg, Inc.
; APPLICANT: Tang, Y. Tom et al
; TITLE OF INVENTION: No. US20040053245A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-030
; CURRENT APPLICATION NUMBER: US/10/276,774
; CURRENT FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 2700
; SOFTWARE: Custom
; SEQ ID NO 1798
; LENGTH: 851
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-276-774-1798

Query Match 98.5%; Score 4614; DB 15; Length 851;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 843; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 7 RKGGGGPKDFGAGIKYNSRHEKVNGLGEVFLPVNNVKKVEKKGPRWVTLAAVILIGLL 66
Db 3 RKGGGGPKDFGAGIKYNSRHEKVNGLGEVFLPVNNVKKVEKKGPRWVTLAAVILIGLL 62
Qy 67 LVLLIGIFLVWHLQYRDVRYQKXNGYMRITNENFVDAYENSNSTEFVSLASKVQDALKL 126
Db 67 LVLLIGIFLVWHLQYRDVRYQKXNGYMRITNENFVDAYENSNSTEFVSLASKVQDALKL 122

QY 127 LYGVPFLGPHYKESAVTAASEGVIAAYWSEPSIPOHLYEAEARVMAEERVMPPRAR 186
DB 123 LYGVPFLGPHYKESAVTAASEGVIAAYWSEPSIPOHLYEAEARVMAEERVMPPRAR 182
QY 187 SLKSFVVTSVVAFPTDSKTQVQRTQDNCSPGLHARGVELMRFTTPGPDSPYPAHARQW 246
DB 183 SLKSFVVTSVVAFPTDSKTQVQRTQDNCSPGLHARGVELMRFTTPGPDSPYPAHARQW 242
QY 247 ALRGDADSVLSTFRSFDLASCDERGRHLVTYVNTLSPMEPHALVOLCGTYPPSYNTLTFH 306
DB 243 ALRGDADSVLSTFRSFDLASCDERGRHLVTYVNTLSPMEPHALVOLCGTYPPSYNTLTFH 302
QY 307 SSQNVLLITLITINTERHHPGEATFPOLPRMSSCGRLRAOQTFNSPYRGHYPPNIDC 366
DB 303 SSQNVLLITLITINTERHHPGEATFPOLPRMSSCGRLRAOQTFNSPYRGHYPPNIDC 362
QY 367 TANIIEVNNQHVKKRFFKFFYLEPGVPAGTCCPDYVEINGEKTCGERSQFVYVTSNSKIT 426
DB 363 TANIIEVNNQHVKKRFFKFFYLEPGVPAGTCCPDYVEINGEKTCGERSQFVYVTSNSKIT 422
QY 427 VRFHSDQSYTDTGFLAELYSYDSSDPCPGQFTCRGRCIRKELRCGMDCTDHSDELNC 486
DB 423 VRFHSDQSYTDTGFLAELYSYDSSDPCPGQFTCRGRCIRKELRCGMDCTDHSDELNC 482
QY 487 SCAGHGFCTCKNFKCKPLFWVCDSDVNDGNSDEGSCSPAQTFRCNSGKCLSKSQCCNG 546
DB 483 SCAGHGFCTCKNFKCKPLFWVCDSDVNDGNSDEGSCSPAQTFRCNSGKCLSKSQCCNG 542
QY 547 KDDCGSDDEASCPKNNVVTCTKHTYRCNLGLCLSKGNPECDGKEDCSDEKDCDGL 606
DB 543 KDDCGSDDEASCPKNNVVTCTKHTYRCNLGLCLSKGNPECDGKEDCSDEKDCDGL 602
QY 607 RSFTQARVVGGTDADGEMPMQVSLHALGQGHICGASLISPMVLVSAHCYIDDRGFRY 666
DB 603 RSFTQARVVGGTDADGEMPMQVSLHALGQGHICGASLISPMVLVSAHCYIDDRGFRY 662
QY 667 SDPTQWTAFLGLHDQSORSAFGVQERRLKRIISHPPNDFTFYDIALLELEKPAEYSSM 726
DB 663 SDPTQWTAFLGLHDQSORSAFGVQERRLKRIISHPPNDFTFYDIALLELEKPAEYSSM 722
QY 727 VRPCLPDASHVPAGKAIWVTGHTQYGGTALILQKGEIRVINTTCENLLPQOITP 786
DB 723 VRPCLPDASHVPAGKAIWVTGHTQYGGTALILQKGEIRVINTTCENLLPQOITP 782
QY 787 RMMCVGFLSGGVDSGQDSGGLSSVADGRIFGAGVSWGDCAGRNKGVYTRLPFLR 846
DB 783 RMMCVGFLSGGVDSGQDSGGLSSVADGRIFGAGVSWGDCAGRNKGVYTRLPFLR 842
QY 847 DWIKENTGV 855
DB 843 DWIKENTGV 851

RESULT 22
US-10-296-115-1143
; Sequence 1143, Application US/10296115
; Publication No. US20040053248A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq Inc
; TITLE OF INVENTION: No. US20040053248A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 784PCT
; CURRENT APPLICATION NUMBER: US/10/296,115
; PRIOR APPLICATION NUMBER: 2002-11-18
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US09/552,317
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 1478
; SEQ ID NO 1143
; LENGTH: 851
; TYPE: PRT
; ORGANISM: Homo sapiens

US-10-296-115-1143

Query Match 98.5%; Score 4614; DB 15; Length 851;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 843; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
QY 7 RKGGGPKDGAALKTNSRHEKYNGLSEGEFLPVANNVKKVEKGGRWVLAAILGLL 66
DB 3 RKGGGPKDGAALKTNSRHEKYNGLSEGEFLPVANNVKKVEKGGRWVLAAILGLL 62
QY 67 LVLLIGFLVWMLQYRDVAVQKRVGYMRTINENPVDAENSNSTFVSLASKVDAKL 126
DB 63 LVLLIGFLVWMLQYRDVAVQKRVGYMRTINENPVDAENSNSTFVSLASKVDAKL 122
QY 127 LYGVPFLGPHYKESAVTAASEGVIAAYWSEPSIPOHLYEAEARVMAEERVMPPRAR 186
DB 123 LYGVPFLGPHYKESAVTAASEGVIAAYWSEPSIPOHLYEAEARVMAEERVMPPRAR 182
QY 187 SLKSFVVTSVVAFPTDSKTQVQRTQDNCSPGLHARGVELMRFTTPGPDSPYPAHARQW 246
DB 183 SLKSFVVTSVVAFPTDSKTQVQRTQDNCSPGLHARGVELMRFTTPGPDSPYPAHARQW 242
QY 247 ALRGDADSVLSTFRSFDLASCDERGRHLVTYVNTLSPMEPHALVOLCGTYPPSYNTLTFH 306
DB 243 ALRGDADSVLSTFRSFDLASCDERGRHLVTYVNTLSPMEPHALVOLCGTYPPSYNTLTFH 302
QY 307 SSQNVLLITLITINTERHHPGEATFPOLPRMSSCGRLRAOQTFNSPYRGHYPPNIDC 366
DB 303 SSQNVLLITLITINTERHHPGEATFPOLPRMSSCGRLRAOQTFNSPYRGHYPPNIDC 362
QY 367 TANIIEVNNQHVKKRFFKFFYLEPGVPAGTCCPDYVEINGEKTCGERSQFVYVTSNSKIT 426
DB 363 TANIIEVNNQHVKKRFFKFFYLEPGVPAGTCCPDYVEINGEKTCGERSQFVYVTSNSKIT 422
QY 427 VRFHSDQSYTDTGFLAELYSYDSSDPCPGQFTCRGRCIRKELRCGMDCTDHSDELNC 486
DB 423 VRFHSDQSYTDTGFLAELYSYDSSDPCPGQFTCRGRCIRKELRCGMDCTDHSDELNC 482
QY 487 SCAGHGFCTCKNFKCKPLFWVCDSDVNDGNSDEGSCSPAQTFRCNSGKCLSKSQCCNG 546
DB 483 SCAGHGFCTCKNFKCKPLFWVCDSDVNDGNSDEGSCSPAQTFRCNSGKCLSKSQCCNG 542
QY 547 KDDCGSDDEASCPKNNVVTCTKHTYRCNLGLCLSKGNPECDGKEDCSDEKDCDGL 606
DB 543 KDDCGSDDEASCPKNNVVTCTKHTYRCNLGLCLSKGNPECDGKEDCSDEKDCDGL 602
QY 607 RSFTQARVVGGTDADGEMPMQVSLHALGQGHICGASLISPMVLVSAHCYIDDRGFRY 666
DB 603 RSFTQARVVGGTDADGEMPMQVSLHALGQGHICGASLISPMVLVSAHCYIDDRGFRY 662
QY 667 SDPTQWTAFLGLHDQSORSAFGVQERRLKRIISHPPNDFTFYDIALLELEKPAEYSSM 726
DB 663 SDPTQWTAFLGLHDQSORSAFGVQERRLKRIISHPPNDFTFYDIALLELEKPAEYSSM 722
QY 727 VRPCLPDASHVPAGKAIWVTGHTQYGGTALILQKGEIRVINTTCENLLPQOITP 786
DB 723 VRPCLPDASHVPAGKAIWVTGHTQYGGTALILQKGEIRVINTTCENLLPQOITP 782
QY 787 RMMCVGFLSGGVDSGQDSGGLSSVADGRIFGAGVSWGDCAGRNKGVYTRLPFLR 846
DB 783 RMMCVGFLSGGVDSGQDSGGLSSVADGRIFGAGVSWGDCAGRNKGVYTRLPFLR 842
QY 847 DWIKENTGV 855
DB 843 DWIKENTGV 851

RESULT 23
US-10-729-807-1
; Sequence 1, Application US/10729807
; Publication No. US20040132158A1
; GENERAL INFORMATION:
; APPLICANT: BANDMAN, Olga; HILLMAN, Jennifer L.


```
/ APPLICANT: TANG, Y. Tom, IAL, Preeti G.
/ APPLICANT: YUE, Henry; AZIMZAI, Yalda
/ APPLICANT: BAUGHN, Mariah R.; LU, Dyung Aina M.
/ TITLE OF INVENTION: HUMAN PEPTIDASES
/ FILE REFERENCE: PR-0651-1 DIV
/ CURRENT APPLICATION NUMBER: US/10/729, 807
/ PRIOR FILING DATE: 2003-12-05
/ PRIOR APPLICATION NUMBER: US 09/889,238
/ PRIOR FILING DATE: 2002-01-24
/ PRIOR APPLICATION NUMBER: PCT/US00/00641
/ PRIOR FILING DATE: 2000-01-11
/ PRIOR APPLICATION NUMBER: US 60/172,247
/ PRIOR FILING DATE: 1999-01-11
/ PRIOR APPLICATION NUMBER: US 60/132,253
/ PRIOR FILING DATE: 1999-05-03
/ PRIOR APPLICATION NUMBER: US 60/136,653
/ PRIOR FILING DATE: 1999-05-27
/ NUMBER OF SEQ ID NOS: 39
/ SOFTWARE: PERL Program
/ SEQ ID NO 1
/ LENGTH: 762
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc feature
/ OTHER INFORMATION: Incyte ID No: 155179CD1
US-10-729-807-1
```

Query Match 89.0%; Score 4167; DB 16; Length 762;

Best Local Similarity 99.6%; Pred. No. 3,6e-310;

Matches 759; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
QY 94 MRTINENFVDAYENSSTEFVSLASKYKDALKLLYSGVPLGPHKESAVTASEGSGVIA 153
D 1 MRTINENFVDAYENSSTEFVSLASKYKDALKLLYSGVPLGPHKESAVTASEGSGVIA 60
QY 154 YWSEFSIPQHLVEAEARVNAEERVVMPPRARSLSKSFVVTSVVAFPTDSKTVQRTDMS 213
D 61 YWSEFSIPQHLVEAEARVNAEERVVMPPRARSLSKSFVVTSVVAFPTDSKTVQRTDMS 120
QY 214 CSFGHARGLVLRFTTTPGFPDSDPYPAHARQNALRDADSVLSLTPRSFDLASCDRGS 273
D 121 CSFGHARGLVLRFTTTPGFPDSDPYPAHARQNALRDADSVLSLTPRSFDLASCDRGS 180
QY 274 DLVTYVNTLSPEMEHALVOLCGTYPSPYNTLPHSSQVNLITLITLTERHHPGEATFPQ 333
D 181 DLVTYVNTLSPEMEHALVOLCGTYPSPYNTLPHSSQVNLITLITLTERHHPGEATFPQ 240
QY 334 LPRMSSCGRLRKAQGTFSNPPYPGHYPNIDCTWNIIEVNNQHVKRFKFFYLLBGPV 393
D 241 LPRMSSCGRLRKAQGTFSNPPYPGHYPNIDCTWNIIEVNNQHVKRFKFFYLLBGPV 300
QY 394 AGTCPPKDYVINGEKYGGERSQFVNTSNKITYRHFHSDSYDTGTGLAETLSDSSDP 453
D 301 AGTCPPKDYVINGEKYGGERSQFVNTSNKITYRHFHSDSYDTGTGLAETLSDSSDP 360
QY 454 PGOFTCTGACIRKEALRCDEMACTDHSDELNCSADGHOFTCKNKKCKPLFWCDSVND 513
D 361 PGOFTCTGACIRKEALRCDEMACTDHSDELNCSADGHOFTCKNKKCKPLFWCDSVND 420
QY 514 CGDNDSDGSCSPAQTFRCNGKCLSKSQCCNGKDCGDSDEASCPKRVNVVTCTKTYR 573
D 421 CGDNDSDGSCSPAQTFRCNGKCLSKSQCCNGKDCGDSDEASCPKRVNVVTCTKTYR 480
QY 574 CLNGCLCTSKNPNBEDGDESDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSD 633
D 481 CLNGCLCTSKNPNBEDGDESDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSD 540
QY 634 ALGCGHICGASLISPNMLVSAAHICYIDRGFRYSDDPTOMTAFGLHDOSORSAAGVOERR 633
D 541 ALGCGHICGASLISPNMLVSAAHICYIDRGFRYSDDPTOMTAFGLHDOSORSAAGVOERR 600
QY 694 LKRIISHPFNDFTFDYDIALLELEKPAEYSSMWVRPICLPDASHVFPAGKAIWVTMGHT 753
```

```
|||||
D 601 LKRIISHPFNDFTFDYDIALLELEKPAEYSSMWVRPICLPDASHVFPAGKAIWVTMGHT 660
QY 754 QYCGTALLIQGEIRVINOCTCENLLPOIIPRMMCVGFLSGVNSCGDSCGPLSSVE 813
D 661 QYCGTALLIQGEIRVINOCTCENLLPOIIPRMMCVGFLSGVNSCGDSCGPLSSVE 720
QY 814 ADGRIFGAGVSWGDCAGRNKPGVYTRPLPRFDMIKENTGV 855
D 721 ADGRIFGAGVSWGDCAGRNKPGVYTRPLPRFDMIKENTGV 762
```

RESULT 24

US-10-097-340-312

/ Sequence 312, Application US/10097340

/ Publication No. US20030087250A1

/ GENERAL INFORMATION:

/ APPLICANT: JOHN MONAHAN

/ APPLICANT: Manjula GANNAVAPAPU

/ APPLICANT: Sebastian HOERSCH

/ APPLICANT: Shubhangi KAMATKAR

/ APPLICANT: Steve G. KOVATS

/ APPLICANT: Rachel E. MEYERS

/ APPLICANT: Michael MORRISBY

/ APPLICANT: Ami SEN

/ APPLICANT: Peter VERIBY

/ APPLICANT: Gordon B. MILLS

/ APPLICANT: Robert C. BAST, Jr.

/ APPLICANT: Karen LU

/ APPLICANT: Rosemarie SCHMANDT

/ APPLICANT: Xumel ZHAO

/ APPLICANT: Karen GALT

/ TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,

/ TITLE OF INVENTION: Assessment, Prevention, and Therapy of Ovarian Cancer

/ FILE REFERENCE: MRI-030

/ CURRENT APPLICATION NUMBER: US/10/097,340

/ PRIOR FILING DATE: 2002-03-14

/ PRIOR APPLICATION NUMBER: 60/276,025

/ PRIOR FILING DATE: 2001-03-14

/ PRIOR APPLICATION NUMBER: 60/325,149

/ PRIOR FILING DATE: 2001-09-26

/ PRIOR APPLICATION NUMBER: 60/276,026

/ PRIOR FILING DATE: 2001-03-14

/ PRIOR APPLICATION NUMBER: 60/324,967

/ PRIOR FILING DATE: 2001/09/26

/ PRIOR APPLICATION NUMBER: 60/311,732

/ PRIOR FILING DATE: 2001-08-10

/ PRIOR APPLICATION NUMBER: 60/325,102

/ PRIOR FILING DATE: 2001-09-26

/ PRIOR APPLICATION NUMBER: 60/323,580

/ PRIOR FILING DATE: 2001-09-19

/ NUMBER OF SEQ ID NOS: 363

/ SOFTWARE: FastSeq for Windows Version 4.0

/ SEQ ID NO 312

/ LENGTH: 782

/ TYPE: PRT

/ ORGANISM: Homo sapiens

US-10-097-340-312

Query Match 88.8%; Score 4158.5; DB 14; Length 782;

Best Local Similarity 90.9%; Pred. No. 1.7e-309;

Matches 777; Conservative 1; Mismatches 4; Indels 73; Gaps 2;

```
QY 1 MGSDBARKKGGGPKDGAAGLKYNSRHEKNGLEEGVEPLPVNNVKKVEKGGPGRWVLLA 60
D 1 MGSDBARKKGGGPKDGAAGLKYNSRHEKNGLEEGVEPLPVNNVKKVEKGGPGRWVLLA 60
QY 61 VILGLLVLLIGIFLVWHLQYRDVRYQVKNQGYMRITNENFVDAYENSSTEFVSLASKY 120
D 61 VILGLLVLLIGIFLVWHLQYRDVRYQVKNQGYMRITNENFVDAYENSSTEFVSLASKY 120
QY 121 KDALKLLYSGVPLGPHKESAVTASEGSGVIAIYWSEFSIPQHLVEAEARVNAEERVVM 180
```

Dh	121	KDMLKLLYSVVPFLGTYHKEASVTAFAESBSVLAAYWSEFSIPQHVLEAEVMAEERVVM	160
Qy	181	LPPRAPSLSKSFVVTSVVAEPFDSKVQRTQDNCSFGJLHANGVELMRFTTGPFPDSYPA	240
Dh	181	LPPRAPSLSKSFVVTSVVAEPFDSKVQRTQDNCSFGJLHANGVELMRFTTGPFPDSYPA	240
Qy	241	HACQOMALRGDADSVLSLTFRSPDASCBERSDILVTVYNTLSPEEPHALVQLGTYPPS	3000
Dh	241	HACQOMALRGDADSVLSLTFRSPDASCBERSDILVTVYNTLSPEEPHALVQLGTYPPS	3000
Qy	301	YNLTFPSSQNVLLITLITWTERRRHPGEATFFQLPRMSSCGGRLLKAKQGTNSPYEGHY	3606
Dh	301	YNLTFPSSQNVLLITLITWTERRRHPGEATFFQLPRMSSCGGRLLKAKQGTNSPYEGHY	3606
Qy	361	PNPIDCTWNIIEVBNNOHVAKRKFYLLBEPVPACTCPXDYVEINBEKXCGERSQFVVT	4200
Dh	361	PNPIDCTWNIIEVBNNOHVAKRKFYLLBEPVPACTCPXDYVEINBEKXCGERSQFVVT	4200
Qy	421	NSNKITVRFPFSDSYNDTDFGLAEYLSYBSSDCPQFCTCRGRCIRKELRCDGADCTDH	4800
Dh	421	NSNKITVRFPFSDSYNDTDFGLAEYLSYBSSDCPQFCTCRGRCIRKELRCDGADCTDH	4800
Qy	481	SDELNCSCDAGHOFCTCKNFKCPLEFVWCDSVNDGDNBDEGCSCPAQTFRCNSGKCLSK	5400
Dh	452	-----RCDAGHOFCTCKNFKCPLEFVWCDSVNDGDNBDEGCSCPAQTFRCNSGKCLSK	4888
Qy	541	SQOQNGKDDCGDSDASCPKVVVVTCTKHTYRCLNGCLSLSGNPECDKEDCSGSDSK	6000
Dh	489	-----NMVVTCTKHTYRCLNGCLSLSGNPECDKEDCSGSDSK	5272
Qy	601	DCOCGLRSPFRQARVGTGDADBGEPWQVSLHALGCGHICGASLISPNMVLVAACIYID	6606
Dh	528	DCOCGLRSPFRQARVGTGDADBGEPWQVSLHALGCGHICGASLISPNMVLVAACIYID	5872
Qy	661	DREPRYSDPQWAFGLHDSORSAPGOEBRLKRIISHPEPNFTFPDYIALLEIEKP	7200
Dh	588	DREPRYSDPQWAFGLHDSORSAPGOEBRLKRIISHPEPNFTFPDYIALLEIEKP	6472
Qy	721	AEVSSWVRPCLPFDASHVFPAGKAIWVTGHTQYGGTGALILQKGEIRVINOTTCEML	7800
Dh	648	AEVSSWVRPCLPFDASHVFPAGKAIWVTGHTQYGGTGALILQKGEIRVINOTTCEML	7072
Qy	781	PQGITPRMNCVGLSGVDSCGDSGGLSVYADGRIFGAGVSWMDGCAGRNKPGVYT	8400
Dh	708	PQGITPRMNCVGLSGVDSCGDSGGLSVYADGRIFGAGVSWMDGCAGRNKPGVYT	7672
Qy	841	RLPLFRDWIKENTGV	855
Dh	768	RLPLFRDWIKENTGV	782
RESULT 25			
US-10-072-012-44			
Sequence 44, Application US/10072012			
Publication No. US20040033493A1			
GENERAL INFORMATION:			
APPLICANT: Tcherev, Vellizar			
APPLICANT: Spytek, Kimberly			
APPLICANT: Zernusen, Bryan			
APPLICANT: Paturajan, Meera			
APPLICANT: Shimkets, Richard			
APPLICANT: Li, Li			
APPLICANT: Gangolli, Neha			
APPLICANT: Pedigaru, Muralidhara			
APPLICANT: Anderson, David W.			
APPLICANT: Rastelli, Luca			
APPLICANT: Miller, Charles E.			
APPLICANT: Gerlach, Valerie			
APPLICANT: Taupier Jr, Raymond J.			
APPLICANT: Gusev, Vladimir Y.			
APPLICANT: Colman, Steven D.			
APPLICANT: Wolenc, Adam R.			

Query Match	87.4%: Score 4092; DB 15; Length 757;	
Best Local Similarity	88.2%: Pred. No. 2e-304;	
Matches 754; Conservative	0; Mismatches 3; Indels 98; Gaps 1	
QY 1 MGSRRARKGCGGPRDFAAGLKYNSRHEKVNGLBEGVEFLPVNNVKKVKGHGGRWVLA 60		
DB 1 MGSRRARKGCGGPRDFAAGLKYNSRHEKVNGLBEGVEFLPVNNVKKVKGHGGRWVLA 60		
QY 61 VLIQLLVLLIGLIGFLVWHLYQYRVRVQVKKVGYRITNENFVDAYENSNSTEFVSLASKY 120		
DB 61 VLIQLLVLLIGLIGFLVWHLYQYRVRVQVKKVGYRITNENFVDAYENSNSTEFVSLASKY 120		
QY 121 KDAKLKLVSGVPLGPRYHKESAVTAFSRGSVIAYYYWSEFSLPHLVLEAEARMYMAERYVM 180		
DB 67 -----LVEAEARMYMAERYVM 82		
QY 181 LPPRRSLKSFVYTSVVAFPDTSKTVORTQDNGSFGILHARGVLEMRFTTGGFPDSPYA 240		
DB 83 LPPRRSLKSFVYTSVVAFPDTSKTVORTQDNGSFGILHARGVLEMRFTTGGFPDSPYA 142		
QY 241 HARCQMLRGDADSVLSTFRSPDLASCDERGSIDLTVYYNTLSPMEPHALVOLCGTYPSS 300		
DB 143 HARCQMLRGDADSVLSTFRSPDLASCDERGSIDLTVYYNTLSPMEPHALVOLCGTYPSS 202		
QY 301 YNLTFHSSQNVLLTLTLNTERRRHGPFAATFPOLPRMSSCGGRLLKAGCTPNSPYYPGHY 360		
DB 203 YNLTFHSSQNVLLTLTLNTERRRHGPFAATFPOLPRMSSCGGRLLKAGCTPNSPYYPGHY 262		
QY 361 PPNIDICTWNIEVPNNQHVKKFKEFYLLLEPVPVPGTCEKDVVEINGEKXCCGRSQFVNTS 420		
DB 263 PPNIDICTWNIEVPNNQHVKKFKEFYLLLEPVPVPGTCEKDVVEINGEKXCCGRSQFVNTS 322		
QY 421 NSKTIIVRFHSDGSYDTDTGLAEYLISYDSDCPQGFCTCRGCTIRKELRCDDGMADCTDH 480		

```
Db 323 NSNKITVRFHSDSYDTGTGLAEYLSYDSSDPGQFTCRGTGRCKIRKELRCDGMADCTDH 382
Qy 481 SDELNGSCDGHQFTCKNFKCKPLFWVCDSDVNDGSDSDGSCSPAOTFRCSNGKCLSK 540
Db 383 SDELNGSCDGHQFTCKNFKCKPLFWVCDSDVNDGSDSDGSCSPAOTFRCSNGKCLSK 442
Qy 541 SQQCGKDCDGDSDSDASCPKRVNVVTCTKHYRCLNGLCLSKGNPECDGKEDCSGSDSEK 600
Db 443 SQQCGKDCDGDSDSDASCPKRVNVVTCTKHYRCLNGLCLSKGNPECDGKEDCSGSDSEK 502
Qy 601 DCCGGLASFTROARVVGCTDADDEGEWPOVSLHALGQGHICGASLISPMNLVSAACICYD 660
Db 503 DCCGGLASFTROARVVGCTDADDEGEWPOVSLHALGQGHICGASLISPMNLVSAACICYD 562
Qy 661 DRGFRYSDPTOMTAFGLHDSQSRAPGVORERLKRILISHPFNDFTFDYDIALLELEXP 720
Db 563 DRGFRYSDPTOMTAFGLHDSQSRAPGVORERLKRILISHPFNDFTFDYDIALLELEXP 622
Qy 721 AEYSSWVRPICLPDASHVPAGKAIWVTGMGHTQYGGTGALLIQKEIRVINOJTTCENLL 780
Db 623 AEYSSWVRPICLPDASHVPAGKAIWVTGMGHTQYGGTGALLIQKEIRVINOJTTCENLL 682
Qy 781 PQQITPRMNCVGLSGVDSCQDSDGSPLSVSEADGRIFQAGVSWGDGCAQRNKPQVYT 840
Db 683 PQQITPRMNCVGLSGVDSCQDSDGSPLSVSEADGRIFQAGVSWGDGCAQRNKPQVYT 742
Qy 841 RLPLFRDWIKENTGV 855
Db 743 RLPLFRDWIKENTGV 757

RESULT 26
US-09-900-751-2
; Sequence 2, Application US/09900751
; Patent No. US20020026653A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Keith D
; APPLICANT: Leventen, Michael W.
; TITLE OF INVENTION: TRANSGENIC MICE CONTAINING SERINE
; TITLE OF INVENTION: PROTEASE GENE DISRUPTIONS
; FILE REFERENCE: R-386
; CURRENT APPLICATION NUMBER: US/09/900,751
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US 60/217,449
; PRIOR FILING DATE: 2000-07-10
; PRIOR APPLICATION NUMBER: US 60/223,170
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: US 60/223,460
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-900-751-2

Query Match 82.9%; Score 3884; DB 9; Length 855;
Best Local Similarity 81.5%; Pred. No. 2e-288;
Matches 697; Conservative 73; Mismatches 85; Indels 0; Gaps 0;
```

```
Qy 181 LPPRASLSKSFVTVSVAPPTDSKTVOQTQDNSCSFGLHARGVELNRFTTPGPSPPYPA 240
Db 181 LPPRASLSKSFVTVSVAPPTDSKTVOQTQDNSCSFGLHARGVELNRFTTPGPSPPYPA 240
Qy 241 HARCQNALRGDADSVLSLFRSPDLASCDERSGDLTVYVNTLSPMFPHALVQCGTYPPS 300
Db 241 HARCQNALRGDADSVLSLFRSPDLASCDERSGDLTVYVNTLSPMFPHALVQCGTYPPS 300
Qy 301 YNLTFFSSQNVLLITLITTERHPGEATFPOLPMSSCGGLRAQGTFFNSPPYPGHY 360
Db 301 YNLTFFSSQNVLLITLITTERHPGEATFPOLPMSSCGGLRAQGTFFNSPPYPGHY 360
Qy 361 PENIDTWNIEVBNQHVNRKFFYLLEPGVAGTCPODYVINEKTYCGERSQFVYTS 420
Db 361 PENIDTWNIEVBNQHVNRKFFYLLEPGVAGTCPODYVINEKTYCGERSQFVYTS 420
Qy 421 NSNKITVRFHSDSYDTGTGLAEYLSYDSSDPGQFTCRGTGRCKIRKELRCDGMADCTDH 480
Db 421 NSNKITVRFHSDSYDTGTGLAEYLSYDSSDPGQFTCRGTGRCKIRKELRCDGMADCTDH 480
Qy 481 SDELNGSCDGHQFTCKNFKCKPLFWVCDSDVNDGSDSDGSCSPAOTFRCSNGKCLSK 540
Db 481 SDELNGSCDGHQFTCKNFKCKPLFWVCDSDVNDGSDSDGSCSPAOTFRCSNGKCLSK 540
Qy 541 SQQCGKDCDGDSDSDASCPKRVNVVTCTKHYRCLNGLCLSKGNPECDGKEDCSGSDSEK 600
Db 541 SQQCGKDCDGDSDSDASCPKRVNVVTCTKHYRCLNGLCLSKGNPECDGKEDCSGSDSEK 600
Qy 601 DCCGGLASFTROARVVGCTDADDEGEWPOVSLHALGQGHICGASLISPMNLVSAACICYD 660
Db 601 DCCGGLASFTROARVVGCTDADDEGEWPOVSLHALGQGHICGASLISPMNLVSAACICYD 660
Qy 661 DRGFRYSDPTOMTAFGLHDSQSRAPGVORERLKRILISHPFNDFTFDYDIALLELEXP 720
Db 661 DRGFRYSDPTOMTAFGLHDSQSRAPGVORERLKRILISHPFNDFTFDYDIALLELEXP 720
Qy 721 AEYSSWVRPICLPDASHVPAGKAIWVTGMGHTQYGGTGALLIQKEIRVINOJTTCENLL 780
Db 721 AEYSSWVRPICLPDASHVPAGKAIWVTGMGHTQYGGTGALLIQKEIRVINOJTTCENLL 780
Qy 781 PQQITPRMNCVGLSGVDSCQDSDGSPLSVSEADGRIFQAGVSWGDGCAQRNKPQVYT 840
Db 781 PQQITPRMNCVGLSGVDSCQDSDGSPLSVSEADGRIFQAGVSWGDGCAQRNKPQVYT 840
Qy 841 RLPLFRDWIKENTGV 855
Db 841 RLPLFRDWIKENTGV 855

RESULT 27
US-10-072-012-355
; Sequence 355, Application US/10072012
; Patent No. US20040003493A1
; GENERAL INFORMATION:
; APPLICANT: Tchervnev, Velizar
; APPLICANT: Spytek, Kimberly
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Patuturajan, Meera
; APPLICANT: Shinkete, Richard
; APPLICANT: Li, Li
; APPLICANT: Gangolli, Baha
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Anderson, David W.
; APPLICANT: Rastelli, Luca
; APPLICANT: Miller, Charles E.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Colman, Steven D.
; APPLICANT: Wolenc, Adam R.
; APPLICANT: Pena, Carol E. A
; APPLICANT: Furtak, Katarzyna
```

```

: APPLICANT: Grosee, William M.
: APPLICANT: Alsobrook II, John P.
: APPLICANT: Lepley, Denise M.
: APPLICANT: Rieger, Daniel K.
: APPLICANT: Burgees, Catherine E.
: TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
: FILE REFERENCE: 21402-258
: CURRENT APPLICATION NUMBER: US/10/072,012
: CURRENT FILING DATE: 2002-01-31
: PRIOR APPLICATION NUMBER: 60/265,102
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: 60/265,514
: PRIOR FILING DATE: 2001-01-31
: PRIOR APPLICATION NUMBER: 60/265,517
: PRIOR FILING DATE: 2001-01-31
: PRIOR APPLICATION NUMBER: 60/265,412
: PRIOR FILING DATE: 2001-01-31
: PRIOR APPLICATION NUMBER: 60/265,395
: PRIOR FILING DATE: 2001-01-31
: PRIOR APPLICATION NUMBER: 60/266,406
: PRIOR FILING DATE: 2001-02-02
: PRIOR APPLICATION NUMBER: 60/266,767
: PRIOR FILING DATE: 2001-02-05
: PRIOR APPLICATION NUMBER: 60/267,057
: PRIOR FILING DATE: 2001-02-07
: PRIOR APPLICATION NUMBER: 60/266,975
: PRIOR FILING DATE: 2001-02-07
: PRIOR APPLICATION NUMBER: 60/267,459
: PRIOR FILING DATE: 2001-02-08
: Remaining Prior Application data removed - See File Wrapper or PALM.
: NUMBER OF SEQ ID NOS: 1391
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 355
: LENGTH: 855
: TYPE: PRT
: ORGANISM: Mus musculus
: US-10-072-012-355

Query Match      82.9%; Score 3884; DB 15; Length 855;
Best Local Similarity 81.5%; Pred. No. 2e-288;
Matches 697; Conservative 73; Mismatches 85; Indels 0; Gaps 0;

```

```

: APPLICANT: Grosee, William M.
: APPLICANT: Alsobrook II, John P.
: APPLICANT: Lepley, Denise M.
: APPLICANT: Rieger, Daniel K.
: APPLICANT: Burgees, Catherine E.
: TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
: FILE REFERENCE: 21402-258
: CURRENT APPLICATION NUMBER: US/10/072,012
: CURRENT FILING DATE: 2002-01-31
: PRIOR APPLICATION NUMBER: 60/265,102
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: 60/265,514
: PRIOR FILING DATE: 2001-01-31
: PRIOR APPLICATION NUMBER: 60/265,517
: PRIOR FILING DATE: 2001-01-31
: PRIOR APPLICATION NUMBER: 60/265,412
: PRIOR FILING DATE: 2001-01-31
: PRIOR APPLICATION NUMBER: 60/265,395
: PRIOR FILING DATE: 2001-01-31
: PRIOR APPLICATION NUMBER: 60/266,406
: PRIOR FILING DATE: 2001-02-02
: PRIOR APPLICATION NUMBER: 60/266,767
: PRIOR FILING DATE: 2001-02-05
: PRIOR APPLICATION NUMBER: 60/267,057
: PRIOR FILING DATE: 2001-02-07
: PRIOR APPLICATION NUMBER: 60/266,975
: PRIOR FILING DATE: 2001-02-07
: PRIOR APPLICATION NUMBER: 60/267,459
: PRIOR FILING DATE: 2001-02-08
: Remaining Prior Application data removed - See File Wrapper or PALM.
: NUMBER OF SEQ ID NOS: 1391
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 355
: LENGTH: 855
: TYPE: PRT
: ORGANISM: Mus musculus
: US-10-072-012-355

Query Match      82.9%; Score 3884; DB 15; Length 855;
Best Local Similarity 81.5%; Pred. No. 2e-288;
Matches 697; Conservative 73; Mismatches 85; Indels 0; Gaps 0;

```

```

; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/266,767
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/267,057
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/266,975
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/267,459
; PRIOR FILING DATE: 2001-02-08
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1391
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 413
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-072-012-413

Query Match      82.6%; Score 3884; DB 15; Length 855;
Best Local Similarity 81.5%; Pred. No. 2e-288;
Matches 697; Conservative 73; Mismatches 85; Indels 0; Gaps 0;

Qy 1 MGSDBARKGGGKPDGAGLKYNSRHKXNGLBEGVFLVNNVKKYKKGPGRWVYLA 60
Db 1 MGSNRGRKAGSGQDFGAGLKYNSRLNNMGFEGBVFLPANNAKYKKGPRRWVYLA 60
Qy 61 VLIGLLVLIGLIGFLVHLYRDVRYQKYGKGYRITNENFVDAYENSNGTEFVSLASKV 120
Db 61 VLFSPLLSLMAGLLVHMFYRNVRYQKVGNGHRIKINELFLDAYENSTSTEFISLASQV 120
Qy 121 KDLAKLIYSGVPLGIPYHKESAVTAFSEGSVIAYYSEFSI POHLVEAEERVMAEERVVM 180
Db 121 KEALKLILYNEVPLVGPYHKESAVTAFSEGSVIAYYSEFSI PPHLAEVDRAMAAVERVVT 180
Qy 181 LPPRARSLSKFVYTSVVAFPDTSKTVORTDONSCEFLHARGVELMFTTGPSPDSYPA 240
Db 181 LPPRARSLSKFVYTSVVAFPDTSKTVORTDONSCEFLHARGVELMFTTGPSPDSYPA 240
Qy 241 HARCOWALRGDADSVLSLTFRSPDLASCDERGSDLVTVNTLSMEPHALVOLCGTYPPS 300
Db 241 HARCOWALRGDADSVLSLTFRSPDLASCDERGSDLVTVNTLSMEPHALVOLCGTYPPS 300
Qy 301 YNLTFFSSQVNLITLITNTERRHGEGEATFFOLPRMSSCGGRIRKAQGTFFNSPYPGHY 360
Db 301 YNLTFFSSQVNLITLITNTERRHGEGEATFFOLPRMSSCGGRIRKAQGTFFNSPYPGHY 360
Qy 361 PPNIDCWMNIEVNNOVKVRPFKFFVLLBEGVAPGCTPKDYVETNGKYGESGQFVYTS 420
Db 361 PPNIDCWMNIEVNNOVKVRPFKFFVLLBEGVAPGCTPKDYVETNGKYGESGQFVYTS 420
Qy 421 NSNKTIVRFHSDQSYTDGTGLAEYLSYSSDPCPGQFTCRGTGR CIRKELRCDGADCTDH 480
Db 421 NSNKTIVRFHSDQSYTDGTGLAEYLSYSSDPCPGQFTCRGTGR CIRKELRCDGADCTDH 480
Qy 481 SDELINSCDAGHOFCKNKKCKPLFWYCDVNDGCGNSDQGSQCPAQTRCSNGKLSK 540
Db 481 SDELINSCDAGHOFCKNKKCKPLFWYCDVNDGCGNSDQGSQCPAQTRCSNGKLSK 540
Qy 541 SDRRYRCNATHTQFTCKNQCFCKPLFWYCDVNDGCGNSDQGSQCPAGSFKCSGKCLPQ 540
Db 541 SDRRYRCNATHTQFTCKNQCFCKPLFWYCDVNDGCGNSDQGSQCPAGSFKCSGKCLPQ 540
Qy 541 SQQCNGKDDGSDSDEASCPKAVVVTCTKHTYRCLNGLCLSKNPECDEKDESDGSD 600
Db 541 SQQCNGKDDGSDSDEASCPKAVVVTCTKHTYRCLNGLCLSKNPECDEKDESDGSD 600
Qy 601 DDCGGLSFTROARVVGTDADDEGEPMQVSLHALGQHLGASLSPNNVLVSAHAYCID 660
Db 601 DDCGGLSFTROARVVGTDADDEGEPMQVSLHALGQHLGASLSPNNVLVSAHAYCID 660
Qy 661 DRGFRYSDPQTWTAFLGLHDSQSRAPGVQERIKRIISHPFNDFFDYDIALLEKRP 720
Db 661 DRGFRYSDPQTWTAFLGLHDSQSRAPGVQERIKRIISHPFNDFFDYDIALLEKRP 720
Qy 721 AEXSSMWRLPCLPDASHVFPAGKAIWYTGNGHNYGSGTGLLILOKGIRIVINOTTCENL 780
Db 721 VETSTVVRPCLPDASHVFPAGKAIWYTGNGHNYGSGTGLLILOKGIRIVINOTTCENL 780
```

```

Qy 781 PQOITPRMVCVGLSGGVDSQCGDSGAPLSVEADGRIFGAGVYVSWGDCGAGRNKGVYT 840
Db 781 PQOITPRMVCVGLSGGVDSQCGDSGAPLSVEADGRIFGAGVYVSWGDCGAGRNKGVYT 840
Qy 841 RLPLFRDWMIXENTGV 855
Db 841 RLPLFRDWMIXENTGV 855

RESULT 29
US-10-072-012-356
; Sequence 356, Application US/10072012
; Publication No. US2004003493A1
; GENERAL INFORMATION:
; APPLICANT: Tchernev, Velizar
; APPLICANT: Spytek, Kimberly
; APPLICANT: Zernusen, Bryan
; APPLICANT: Patuturajan, Meera
; APPLICANT: Shinkets, Richard
; APPLICANT: Li, Li
; APPLICANT: Gangolli, Baha
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Anderson, David W.
; APPLICANT: Rastelli, Luca
; APPLICANT: Miller, Charles E.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Colman, Steven D.
; APPLICANT: Molenc, Adam R.
; APPLICANT: Pena, Carol E. A
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Grosse, William M.
; APPLICANT: Alsobrook II, John P.
; APPLICANT: Lepley, Denise M.
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Burgess, Catherine E.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-258
; CURRENT APPLICATION NUMBER: US/10/072,012
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 60/265,102
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/265,514
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,412
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,395
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/266,406
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/266,767
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/267,057
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/266,975
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/267,459
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1391
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 356
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-10-072-012-356

Query Match      82.6%; Score 3866; DB 15; Length 855;
Best Local Similarity 80.8%; Pred. No. 4.7e-287;
```

```
Matches 691; Conservative 79; Mismatches 85; Indels 0; Gaps 0;
QY 1 MSDDRARRKGGGGPDPGAGLTKNSRHKVNGLEGVYFLPNNVKKYKPKGKRWVYLA 60
DB 1 MGNNGRKGAGGGGDDFGAGLTKNSRLNNMGFEVGFLLPNNAKQYKKGPRRWVYVA 60
QY 61 VLLGLLVLVLIGLGLVWHLQYRDVRYQKXNGYMRITNENFVDAYESNSTEFLVSLASKV 120
DB 61 VVFSFLLSLMAGLLVWHLVHFRVNRVYQKXNGYMRITNENFVDAYESNSTEFLVSLASKV 120
QY 121 KDALKLKLYSGVPLGPRHKSAYTAFASEGSVIAYYSEFSI PQHLVEARVNAEERVVM 180
DB 121 KEALKLKLYSEVPLGPRHKSAYTAFASEGSVIAYYSEFSI PPHLEBEVDRAVAVERVVT 180
QY 181 LPPRARSLSKFFVYTSVAAPFTDSKTVOPTQDNSSFGLLHARGVLMRFTTGPFPDSYPA 240
DB 181 LPPRARSLSKFFVYTSVAAPFTDSKTVOPTQDNSSFGLLHARGVLMRFTTGPFPDSYPA 240
QY 241 HARGQNALRGDADSVSLTFRSPDLASCDERGSDDLVTYNTLSPMERHALVOLCGTYPPS 300
DB 241 HARGQNALRGDADSVSLTFRSPDLASCDERGSDDLVTYNTLSPMERHALVOLCGTYPPS 300
QY 301 YNLTFHSSQNVLLTLITNTERHPRGEATFQULPRNMSCGRLRKAQGTFSNPPYRGHY 360
DB 301 YNLTFHSSQNVLLTLITNTERHPRGEATFQULPRNMSCGRLRKAQGTFSNPPYRGHY 360
QY 361 PPNIDCTWNIENPVNNOYKTRFEKFFYLLEPGVPAGTCKPYVEINSGKYCGERSQFVYTS 420
DB 361 PPNINCTWNIENPVNNOYKTRFEKFFYLLEPGVPAGTCKPYVEINSGKYCGERSQFVYTS 420
QY 421 NSNKTITRPHSDQSYTGTGFLAEYLSYDSDPGQFTCRGTGRCIRLELRCDACTDH 480
DB 421 NSNKTITRPHSDQSYTGTGFLAEYLSYDSDPGQFTCRGTGRCIRLELRCDACTDH 480
QY 481 SDELINCSGADGHOFTCKNKECKPLFWYCDVNDGSDNSDEGSCSPAQTRFCSNGKCLSK 540
DB 481 SDEHRCGCMATHOQWCMGNOFCPLFWYCDVNDGSDNSDEGSCSPAQTRFCSNGKCLSK 540
QY 541 SQQCGNGDDGSDGDEASCPKNNVYVCTKATYRCLNGLCLSKNPPEDDGKEDSGDSDGDEK 600
DB 541 SQQCGNGDDGSDGDEASCPKNNVYVCTKATYRCLNGLCLSKNPPEDDGKEDSGDSDGDEK 600
QY 601 DCCGGLSFTROARVVGSTDADEGEPMOYSLHLAGGHLGASLISPMNLVLSAAHCYID 660
DB 601 NCCDGLSFTROARVVGSTDADEGEPMOYSLHLAGGHLGASLISPMNLVLSAAHCYID 660
QY 661 DRGFRYSPTQMTAFGLGLHDSQSRAPGVQERRLKRITISHPFNDFTFDYDIALLELEKP 720
DB 661 ETIFKYSDHMTMTAFGLGLHDSQSRAPGVQERRLKRITISHPFNDFTFDYDIALLELEKP 720
QY 721 AEYSWVRPICLPDASHVFPAGKAIWYTGHTOYGTGALLQKGBIRVYNQTTGCEML 780
DB 721 AEYSTVVRPICLPDASHVFPAGKAIWYTGHTOYGTGALLQKGBIRVYNQTTGCEML 780
QY 781 PQOITPRMWCYGLSGGVSDSCQSGGPGLSVEADGRIFGAGVYVSMGDCGGRKPGVYT 840
DB 781 PQOITPRMWCYGLSGGVSDSCQSGGPGLSVEADGRIFGAGVYVSMGDCGGRKPGVYT 840
QY 841 RLPLFRDMIKENTGV 855
DB 841 RLPFRDMIKENTGV 855
RESULT 30
US-10-072-012-414
; Sequence 414, Application US/10072012
; Publication No. US20040033493A1
; GENERAL INFORMATION:
; APPLICANT: Ichnerev, Velizar
; APPLICANT: Zeyneen, Bryan
; APPLICANT: Pacurajan, Meera
; APPLICANT: Shimkets, Richard
```

```
APPLICANT: Li, Li
APPLICANT: Gangoli, Baha
APPLICANT: Padigaru, Murajidhara
APPLICANT: Anderson, David W.
APPLICANT: Rastelli, Luca
APPLICANT: Miller, Charles E.
APPLICANT: Gerlach, Valerie
APPLICANT: Taupier Jr, Raymond J.
APPLICANT: Gusev, Vladimir Y.
APPLICANT: Coleman, Steven D.
APPLICANT: Wolenc, Adam R.
APPLICANT: Pena, Carol E. A
APPLICANT: Furtak, Katarzyna
APPLICANT: Grose, William M.
APPLICANT: Alsbrook II, John P.
APPLICANT: Lepley, Denise M.
APPLICANT: Rieger, Daniel K.
APPLICANT: Burgess, Catherine E.
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
FILE REFERENCE: 21402-258
CURRENT APPLICATION NUMBER: US/10/072,012
CURRENT FILING DATE: 2002-01-31
PRIOR APPLICATION NUMBER: 60/265,102
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: 60/265,514
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265,517
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265,412
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265,395
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/266,406
PRIOR FILING DATE: 2001-02-02
PRIOR APPLICATION NUMBER: 60/266,767
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: 60/267,057
PRIOR FILING DATE: 2001-02-07
PRIOR APPLICATION NUMBER: 60/266,975
PRIOR FILING DATE: 2001-02-07
PRIOR APPLICATION NUMBER: 60/267,459
PRIOR FILING DATE: 2001-02-08
NUMBER OF SEQ ID NOS: 1391
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 414
LENGTH: 855
TYPE: PRT
ORGANISM: Rattus norvegicus
US-10-072-012-414
Query Match 82.6%; Score 3866; DB 15; Length 855;
Best Local Similarity 80.8%; Pred. No. 4.7e-287;
Matches 691; Conservative 79; Mismatches 85; Indels 0; Gaps 0;
QY 1 MSDDRARRKGGGGPDPGAGLTKNSRHKVNGLEGVYFLPNNVKKYKPKGKRWVYLA 60
DB 1 MGNNGRKGAGGGGDDFGAGLTKNSRLNNMGFEVGFLLPNNAKQYKKGPRRWVYVA 60
QY 61 VLLGLLVLVLIGLGLVWHLQYRDVRYQKXNGYMRITNENFVDAYESNSTEFLVSLASKV 120
DB 61 VVFSFLLSLMAGLLVWHLVHFRVNRVYQKXNGYMRITNENFVDAYESNSTEFLVSLASKV 120
QY 121 KDALKLKLYSGVPLGPRHKSAYTAFASEGSVIAYYSEFSI PQHLVEARVNAEERVVM 180
DB 121 KEALKLKLYSEVPLGPRHKSAYTAFASEGSVIAYYSEFSI PPHLEBEVDRAVAVERVVT 180
QY 181 LPPRARSLSKFFVYTSVAAPFTDSKTVOPTQDNSSFGLLHARGVLMRFTTGPFPDSYPA 240
DB 181 LPPRARSLSKFFVYTSVAAPFTDSKTVOPTQDNSSFGLLHARGVLMRFTTGPFPDSYPA 240
QY 241 HARGQNALRGDADSVSLTFRSPDLASCDERGSDDLVTYNTLSPMERHALVOLCGTYPPS 300
DB 241 HARGQNALRGDADSVSLTFRSPDLASCDERGSDDLVTYNTLSPMERHALVOLCGTYPPS 300
```

```

Db      241 HARCQWVLRGDADSVLSLTRSFVAPCDGHDSDLVTVYDLSLSPMEPHAVVRLCGTSPS 300
QY      301 YNLTFHSSQNVLLITLLINTERRHGPGEATFFQLPRMSSCGRLRKAQGTENSPYPGNY 360
Db      301 YNLTFHSSQNVFLVTLITNDRHRHGPGEATFFQLPRMSSCGRLRKAQGTENSPYPGNY 360
QY      361 PPNIDCTWNIIEVNNQVAKVRFKFFYLIEPGVPAGTCPKDYVEINGEKYCGERSQFVYTS 420
Db      361 PPNINCTWNIKVNNRVKVRFKLFYLVDPNI PVGSC TKDYVEINGEKFCGERSQFVYSS 420
QY      421 NSNKITVRFHSDOSYTDGTGFLAEYLSYDSDPCPGQFTCRTRGRCIRKELRCDGWADCTDH 480
Db      421 NSSKITVHFHSDHSHYTTDTGFLAEYLSYDSDPCPGMFMCKTGRCIRKDLRCDGWADCPDY 480
QY      481 SDELNCSGDAAGHOFCKNFKCKPLFWVCDSVNDCGNSDEQSCSPAQTFRCNSGKCLSK 540
Db      481 SDEHRCRCNATHQFMCKNQFCCKPLFWVCDSVNDCGDSDEEGSCSPAQSFKCSNGKCLPQ 540
QY      541 SQQNGKDDCGDSDASCPKRVNVYCTKATYRCLNGLCLSKGNPECDGKEDCGSDSEK 600
Db      541 SQQNGKDDCGDSDASCDNVAVNSCTKYTYRCONGICLNKGNPECDGKEDCGSDSEK 600
QY      601 DCDGLRSFTROARVVGTDADEGEMPVOYSLHALGQGHICGASLISPNMLVSAHGYID 660
Db      601 NCDGLRSFTKQARVVGSTNADEGEMPVOYSLHALGQGHICGASLISPDMLVSAHCFOD 660
QY      661 DRGFRYSDPQTQWTAFLGLHDQSORAPVOERBLKRIISHPFNDFTFYDIALLELEKP 720
Db      661 ETIFKYSDHMTWTFGLLDQSKRSASGVQEHKLRITTHPSFNDFTFYDIALLELEKP 720
QY      721 AEYSMWRFPICLPDASHVFPAGKAIWYTGNGHTOYGTGALILQKGEIRVINOTTGENTL 780
Db      721 AEYSTVVRPICLPNTVHVPAGKAIWYTGNGHTKEGGTGALILQKGEIRVINOTTCELL 780
QY      781 PQQITPRMFCVGLSGVDSCQSGDGGPLSSVEADGRIFGAGVYVWGDGCGAGRKPGVYT 840
Db      781 PQQITPRMFCVGLSGVDSCQSGDGGPLSSVEKDGRIFGAGVYVWGECCAGRKPGVYT 840
QY      841 RLPLFRDWMIKENTGV 855
Db      841 RIPEVRDWMIKEQTV 855

```

Search completed: November 29, 2004, 08:49:57
 Job time : 118.183 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 29, 2004, 08:20:59 : Search time 32.2432 Seconds
(without alignments)
1758.570 Million cell updates/sec

Title: US09-936-333-27

Perfect score: 4683
Sequence: 1 MGSDRARKGGCGPDPGAGL.....RGVYTRLPFRDWIKENTGV 855

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 6618000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 100 summaries

Database :

1: /cgnt2_6/ptodata/1/1aa/5A.COMB.pdp:*
2: /cgnt2_6/ptodata/1/1aa/5B.COMB.pdp:*
3: /cgnt2_6/ptodata/1/1aa/6A.COMB.pdp:*
4: /cgnt2_6/ptodata/1/1aa/6B.COMB.pdp:*
5: /cgnt2_6/ptodata/1/1aa/PCTUS.COMB.pdp:*
6: /cgnt2_6/ptodata/1/1aa/backfile1.pdp:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4653	99.4	855	2	US-09-027-337-2
2	4653	99.4	855	4	US-09-644-600-2
3	4653	99.4	855	4	US-09-654-600A-2
4	3793	81.0	902	4	US-09-644-600-10
5	3793	81.0	902	4	US-09-654-600A-10
6	1305	27.9	241	4	US-09-657-986B-2
7	700.5	15.0	798	1	US-08-200-900A-2
8	700.5	15.0	798	1	US-08-794-042-2
9	700.5	15.0	798	5	PCT-US94-00616-2
10	583	12.4	407	3	US-09-734-675-4
11	557.5	11.9	492	4	US-09-685-166A-895
12	557.5	11.9	492	4	US-09-879-792-14
13	557.5	11.9	492	4	US-09-679-426-895
14	555.5	11.9	492	4	US-09-342-749-2
15	555.5	11.9	492	4	US-09-691-840-2
16	543.5	11.6	235	3	US-08-944-483-65
17	540.5	11.5	235	3	US-08-807-151-3
18	540.5	11.5	235	3	US-09-478-957-3
19	536.5	11.1	454	4	US-09-518-046-2
20	518	11.1	446	4	US-10-177-661-4
21	517.5	11.1	418	4	US-10-177-661-6
22	509.5	10.9	283	3	US-08-807-151-1
23	509.5	10.9	283	3	US-09-478-957-1
24	509.5	10.9	477	4	US-10-177-661-2
25	509.5	10.9	562	4	US-09-879-792-12
26	508.5	10.9	418	1	US-08-508-448C-25
27	508.5	10.9	418	4	US-09-370-838-82

28	508.5	10.9	418	4	US-09-370-838-83	Sequence 83, Appl
29	508.5	10.9	418	4	US-09-854-133-82	Sequence 82, Appl
30	508.5	10.9	418	4	US-09-854-133-83	Sequence 83, Appl
31	507.5	10.8	418	4	US-09-370-838-62	Sequence 62, Appl
32	507.5	10.8	418	4	US-09-854-133-62	Sequence 62, Appl
33	507.5	10.8	638	2	US-08-681-151-3	Sequence 3, Appl
34	506.5	10.8	790	4	US-08-991-761A-13	Sequence 13, Appl
35	501.5	10.7	376	4	US-09-820-002-2	Sequence 2, Appl
36	494.5	10.6	232	4	US-08-508-448C-19	Sequence 19, Appl
37	492	10.5	416	2	US-09-000-466-2	Sequence 2, Appl
38	491.5	10.5	546	6	5200340-6	Sequence 2, Appl
39	489	10.4	791	1	US-08-643-219-1	Sequence 1, Appl
40	489	10.4	791	3	US-08-851-350-1	Sequence 1, Appl
41	488.5	10.4	583	4	US-09-976-594-837	Sequence 837, App
42	488	10.4	417	4	US-09-820-002-4	Sequence 4, Appl
43	487	10.4	791	2	US-09-131-995-1	Sequence 1, Appl
44	487	10.4	791	2	US-08-832-087B-1	Sequence 1, Appl
45	487	10.4	791	3	US-09-132-154-1	Sequence 1, Appl
46	487	10.4	791	4	US-08-991-761A-6	Sequence 6, Appl
47	487	10.4	791	4	US-08-924-287A-1	Sequence 1, Appl
48	487	10.4	810	1	US-07-854-603-2	Sequence 2, Appl
49	487	10.4	810	1	US-08-147-000B-29	Sequence 29, Appl
50	487	10.4	810	3	US-09-086-514-1	Sequence 1, Appl
51	487	10.4	810	4	US-09-192-012-5	Sequence 5, Appl
52	487	10.4	810	4	US-09-403-736-1	Sequence 1, Appl
53	487	10.4	810	4	US-09-701-265-1	Sequence 1, Appl
54	487	10.4	810	6	5200340-8	Sequence 1, Appl
55	485.5	10.4	812	4	US-08-991-761A-7	Sequence 7, Appl
56	484.5	10.3	790	1	US-08-469-486-54	Sequence 54, Appl
57	484.5	10.3	790	2	US-08-469-558-54	Sequence 54, Appl
58	484	10.3	814	1	US-08-750-711-1	Sequence 1, Appl
59	483.5	10.3	812	1	US-08-248-629A-1	Sequence 1, Appl
60	483.5	10.3	812	1	US-08-451-932-1	Sequence 1, Appl
61	483.5	10.3	812	1	US-08-452-260-1	Sequence 1, Appl
62	483.5	10.3	812	1	US-08-326-785-1	Sequence 1, Appl
63	483.5	10.3	812	2	US-08-612-788-1	Sequence 1, Appl
64	483.5	10.3	812	2	US-08-605-598B-1	Sequence 1, Appl
65	483.5	10.3	812	2	US-08-429-743-1	Sequence 1, Appl
66	483.5	10.3	812	2	US-08-866-735-1	Sequence 1, Appl
67	483.5	10.3	812	2	US-09-066-028-1	Sequence 1, Appl
68	483.5	10.3	812	4	US-09-192-012-3	Sequence 3, Appl
69	483.5	10.3	812	4	US-09-335-325-1	Sequence 1, Appl
70	483.5	10.3	812	4	US-08-991-761A-12	Sequence 12, Appl
71	483.5	10.3	812	5	PCT-US95-05107-1	Sequence 1, Appl
72	479.5	10.2	415	1	US-08-073-531B-1	Sequence 1, Appl
73	479.5	10.2	415	1	US-08-766-288-1	Sequence 1, Appl
74	477	10.2	810	4	US-08-991-761A-11	Sequence 11, Appl
75	475	10.1	328	4	US-09-386-642-11	Sequence 11, Appl
76	474.5	10.1	256	2	US-09-027-337-3	Sequence 3, Appl
77	474.5	10.1	256	4	US-09-644-600-3	Sequence 3, Appl
78	474.5	10.1	256	4	US-09-654-600A-3	Sequence 3, Appl
79	474.5	10.1	415	1	US-08-295-411-2	Sequence 2, Appl
80	474.5	10.1	415	5	US-08-955-411-2	Sequence 2, Appl
81	474.5	10.1	415	5	PCT-US92-10242-2	Sequence 2, Appl
82	474.5	10.1	461	3	US-08-742-877-2	Sequence 2, Appl
83	472	10.1	405	3	US-09-734-675-2	Sequence 2, Appl
84	472	10.1	461	6	5521070-2	Sequence 2, Appl
85	471.5	10.1	315	4	US-09-386-653A-9	Sequence 9, Appl
86	471	10.1	406	4	US-09-851-588-6	Sequence 6, Appl
87	471	10.1	423	4	US-09-656-002-2	Sequence 2, Appl
88	471	10.1	435	3	US-09-008-271A-6	Sequence 6, Appl
89	471	10.1	437	4	US-09-607-745-2	Sequence 2, Appl
90	471	10.1	437	4	US-09-861-588-8	Sequence 8, Appl
91	471	10.1	461	3	US-09-053-811A-21	Sequence 21, Appl
92	471	10.1	461	4	US-10-133-907-5	Sequence 5, Appl
93	469.5	10.0	255	3	US-08-944-483-67	Sequence 67, Appl
94	469	10.0	319	4	US-09-386-642-12	Sequence 12, Appl
95	468.5	10.0	415	4	US-09-118-748-2	Sequence 2, Appl
96	468	10.0	455	3	US-09-261-416-2	Sequence 2, Appl
97	468	10.0	527	6	5520913-1	Sequence 6, Appl
98	466.5	10.0	290	4	US-09-386-653A-7	Sequence 7, Appl
99	466.5	10.0	290	4	US-10-140-002-222	Sequence 222, App
100	463	9.9	527	2	US-08-811-949-39	Sequence 39, Appl

ALIGNMENTS

```
RESULT 1
US-09-027-337-2
; Sequence 2, Application US/09027337B
; Patent No. 5972616
; GENERAL INFORMATION:
; APPLICANT: Tanimoto, Timothy J.
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease Overexpressed in
; FILE REFERENCE: D6064
; CURRENT APPLICATION NUMBER: US/09/027,337B
; NUMBER OF SEQ ID NOS: 13
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Amino acid sequence of TADG-15 encoded by nucleotides
; Patent No. 5972616
US-09-027-337-2

Query Match      99.4%; Score 4653; DB 2; Length 855;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 MGSDRAKGGGGPGDFAGLKYNRSRHKVNGLEGEVFLPVNNVKYKKGPRWVLLAA 60
DB 1 MGSDRAKGGGGPGDFAGLKYNRSRHKVNGLEGEVFLPVNNVKYKKGPRWVLLAA 60
QY 61 VLIGLLVLLIGTFLVHMLQYRDVAVQKXNGYRITNEFVDAYENSNSTEYVLSASKV 120
DB 61 VLIGLLVLLIGTFLVHMLQYRDVAVQKXNGYRITNEFVDAYENSNSTEYVLSASKV 120
QY 121 KDLAKLLYSVGPLGPRYHKSAYTAFASEGSVIAYTSEFSIPQHLVEAEKRWAAEERVVM 180
DB 121 KDLAKLLYSVGPLGPRYHKSAYTAFASEGSVIAYTSEFSIPQHLVEAEKRWAAEERVVM 180
QY 181 LPPRARSLSKSFVVTSVVAFPTDSKTVOQTQDNCSCFGLHARGVELMFTTGPFPDSPYPA 240
DB 181 LPPRARSLSKSFVVTSVVAFPTDSKTVOQTQDNCSCFGLHARGVELMFTTGPFPDSPYPA 240
QY 241 HARCQALRGDADSVSLTFRSPDLASCDERGSDLVTYVNTLSPMERHALVOLCGTYPSP 300
DB 241 HARCQALRGDADSVSLTFRSPDLASCDERGSDLVTYVNTLSPMERHALVOLCGTYPSP 300
QY 301 YNLTFHSSQNVLLITLTINTERRHHPGEATFFQLPRMSSCGRLRKAQGTFFNSPYTGHY 360
DB 301 YNLTFHSSQNVLLITLTINTERRHHPGEATFFQLPRMSSCGRLRKAQGTFFNSPYTGHY 360
QY 361 PPNIIDCTWNIETVNNQHVAVKFEFFYLLEBPVPAATCPKDYVEINGEKYGERSQFVVT 420
DB 361 PPNIIDCTWNIETVNNQHVAVKFEFFYLLEBPVPAATCPKDYVEINGEKYGERSQFVVT 420
QY 421 NSNKITTYRHSDDSYDTGTFLAEYLSYDSDPCPGQGTCTGTGRCIRKELRCDGMACTDH 480
DB 421 NSNKITTYRHSDDSYDTGTFLAEYLSYDSDPCPGQGTCTGTGRCIRKELRCDGMACTDH 480
QY 481 SDELNGSCDAGHOTCTCKNFKCPFLFWVCDSVNDGSDSCGCPQOTFRCSNGKCLSK 540
DB 481 SDELNGSCDAGHOTCTCKNFKCPFLFWVCDSVNDGSDSCGCPQOTFRCSNGKCLSK 540
QY 541 SQOQNGKDDGSDGDEASCPKVVNVVCTKHTYRCLNGLCLSKNPECDGKEDCDSDGDEK 600
DB 541 SQOQNGKDDGSDGDEASCPKVVNVVCTKHTYRCLNGLCLSKNPECDGKEDCDSDGDEK 600
QY 601 DDCGGLRSFTFRQARVVGCTADBEQWVQVSLHALGGGHTCGASLISPNMLVSAAHCTID 660
```

```
DB 601 DDCGGLRSFTFRQARVVGCTADBEQWVQVSLHALGGGHTCGASLISPNMLVSAAHCTID 660
QY 661 DRGFRYSDDPTQMTAPLGIHQDSORSAQVQERLKRILSHPFNDPFDYDILLLEKXP 720
DB 661 DRGFRYSDDPTQMTAPLGIHQDSORSAQVQERLKRILSHPFNDPFDYDILLLEKXP 720
QY 721 AEYSSWVRPICLPDASHVFPAGRAIWTGHTQYGGTGALLQKEIRVINOJTCEML 780
DB 721 AEYSSWVRPICLPDASHVFPAGRAIWTGHTQYGGTGALLQKEIRVINOJTCEML 780
QY 781 PQOITPRMVCVGLSGVSDSCQDSSGFLSSVADGRIFAGVYVSWGDCAGRNKPEVYT 840
DB 781 PQOITPRMVCVGLSGVSDSCQDSSGFLSSVADGRIFAGVYVSWGDCAGRNKPEVYT 840
QY 841 RLPLFRDWMKENTGV 855
DB 841 RLPLFRDWMKENTGV 855
```

RESULT 2

```
US-09-644-600-2
; Sequence 2, Application US/09644600
; Patent No. 6431500
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease
; FILE REFERENCE: D6064CIP/D
; CURRENT APPLICATION NUMBER: US/09/644,600
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: 09/421,213
; PRIOR FILING DATE: 1999-10-20
; PRIOR APPLICATION NUMBER: 09/027,337
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: TADG-15
US-09-644-600-2
```

```
Query Match      99.4%; Score 4653; DB 4; Length 855;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 MGSDRAKGGGGPKDFAGLKYNRSRHKVNGLEGEVFLPVNNVKYKKGPRWVLLAA 60
DB 1 MGSDRAKGGGGPKDFAGLKYNRSRHKVNGLEGEVFLPVNNVKYKKGPRWVLLAA 60
QY 61 VLIGLLVLLIGTFLVHMLQYRDVAVQKXNGYRITNEFVDAYENSNSTEYVLSASKV 120
DB 61 VLIGLLVLLIGTFLVHMLQYRDVAVQKXNGYRITNEFVDAYENSNSTEYVLSASKV 120
QY 121 KDLAKLLYSVGPLGPRYHKSAYTAFASEGSVIAYTSEFSIPQHLVEAEKRWAAEERVVM 180
DB 121 KDLAKLLYSVGPLGPRYHKSAYTAFASEGSVIAYTSEFSIPQHLVEAEKRWAAEERVVM 180
QY 181 LPPRARSLSKSFVVTSVVAFPTDSKTVOQTQDNCSCFGLHARGVELMFTTGPFPDSPYPA 240
DB 181 LPPRARSLSKSFVVTSVVAFPTDSKTVOQTQDNCSCFGLHARGVELMFTTGPFPDSPYPA 240
QY 241 HARCQALRGDADSVSLTFRSPDLASCDERGSDLVTYVNTLSPMERHALVOLCGTYPSP 300
DB 241 HARCQALRGDADSVSLTFRSPDLASCDERGSDLVTYVNTLSPMERHALVOLCGTYPSP 300
QY 301 YNLTFHSSQNVLLITLTINTERRHHPGEATFFQLPRMSSCGRLRKAQGTFFNSPYTGHY 360
DB 301 YNLTFHSSQNVLLITLTINTERRHHPGEATFFQLPRMSSCGRLRKAQGTFFNSPYTGHY 360
```

```
QY 361 PPNIDCTWNIIEVNNQHVKVRFKFYYLLBPGVAGTCPKDYVEINGEKYCGERSQFVYTS 420
DB 361 PPNIDCTWNIIEVNNQHVKVSFKFYYLLBPGVAGTCPKDYVEINGEKYCGERSQFVYTS 420
QY 421 NSNKITVRFHSDOSYTDYTGFLAEYLSYDSSDPGQFTCRTRGRCIRKELRCDDGADCTDH 480
DB 421 NSNKITVRFHSDOSYTDYTGFLAEYLSYDSSDPGQFTCRTRGRCIRKELRCDDGADCTDH 480
QY 481 SDELNCSCDAGHQTCKNFKCKPLFWYCDSDVNDGSDNSDEGSCCPAQTRFCSNGKCLSK 540
DB 481 SDELNCSCDAGHQTCKNFKCKPLFWYCDSDVNDGSDNSDEGSCCPAQTRFCSNGKCLSK 540
QY 541 SQQNGKDDGSDGSDASCPKRVNVYTCIKHYRCLNGLCLSKGNPECDGKEDCSGSDSK 600
DB 541 SQQNGKDDGSDGSDASCPKRVNVYTCIKHYRCLNGLCLSKGNPECDGKEDCSGSDSK 600
QY 601 DCDGGLSFTROARVVGCTADDEGEMPMOYSLHALGCGHICGASLISPMNLVSAHACYID 660
DB 601 DCDGGLSFTROARVVGCTADDEGEMPMOYSLHALGCGHICGASLISPMNLVSAHACYID 660
QY 661 DRGFRYSDDPTQWTAFLGLHDOSQSRAPGVQERRLKRIISHPFNDFTFDYDIALLELEKP 720
DB 661 DRGFRYSDDPTQWTAFLGLHDOSQSRAPGVQERRLKRIISHPFNDFTFDYDIALLELEKP 720
QY 721 AEYSMWRIPICLPASHVFPAGKAIWYTGWHTQYGGTGALILQKGEIRVINTTCENL 780
DB 721 AEYSMWRIPICLPASHVFPAGKAIWYTGWHTQYGGTGALILQKGEIRVINTTCENL 780
QY 781 PQQITPRMCMCVGLSGVDSCQSDGSPSSVEADGRIFGAGVSWGDCAGRNKPGVYT 840
DB 781 PQQITPRMCMCVGLSGVDSCQSDGSPSSVEADGRIFGAGVSWGDCAGRNKPGVYT 840
QY 841 RLPLFRDWMIKENTGV 855
DB 841 RLPLFRDWMIKENTGV 855
```

```
RESULT 3
US-09-654-600A-2
; Sequence 2, Application US/09654600A
; Patent No. 6649741
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Tanimoto, Hirotoshi
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease
; FILE REFERENCE: D6064CIP/D
; CURRENT APPLICATION NUMBER: US/09/654,600A
; CURRENT FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 09/421,213
; 09/027,337
; PRIOR FILING DATE: 1999-10-20
; 1998-02-20
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: TADG-15
US-09-654-600A-2
```

```
Query Match 99.4%; Score 4653; DB 4; Length 855;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 121 KDALKLISGVPLFGPHKESAVTAPSEGSVIAYYWSEFSIPOHLVEAEERVAEERVM 180
DB 121 KDALKLISGVPLFGPHKESAVTAPSEGSVIAYYWSEFSIPOHLVEAEERVAEERVM 180
QY 181 LPPARSLKSFVYTSVVAFPDTSKTYVORTDNSCSFGLHARGVELMFTTGGPDPSPYA 240
DB 181 LPPARSLKSFVYTSVVAFPDTSKTYVORTDNSCSFGLHARGVELMFTTGGPDPSPYA 240
QY 241 HARCQMALRGDADSVLSLTRSFDLASCDERGSGLTYVNTTSLPMEPHALVOLCGTYPSP 300
DB 241 HARCQMALRGDADSVLSLTRSFDLASCDERGSGLTYVNTTSLPMEPHALVOLCGTYPSP 300
QY 301 YNLTFHSSQNVLLITLTINTEERRHPGEATFPOLPRMSCGGRLRKAQGFNSPYYPGHY 360
DB 301 YNLTFHSSQNVLLITLTINTEERRHPGEATFPOLPRMSCGGRLRKAQGFNSPYYPGHY 360
QY 361 PPNIDCTWNIIEVNNQHVKVRFKFYYLLBPGVAGTCPKDYVEINGEKYCGERSQFVYTS 420
DB 361 PPNIDCTWNIIEVNNQHVKVSFKFYYLLBPGVAGTCPKDYVEINGEKYCGERSQFVYTS 420
QY 421 NSNKITVRFHSDOSYTDYTGFLAEYLSYDSSDPGQFTCRTRGRCIRKELRCDDGADCTDH 480
DB 421 NSNKITVRFHSDOSYTDYTGFLAEYLSYDSSDPGQFTCRTRGRCIRKELRCDDGADCTDH 480
QY 481 SDELNCSCDAGHQTCKNFKCKPLFWYCDSDVNDGSDNSDEGSCCPAQTRFCSNGKCLSK 540
DB 481 SDELNCSCDAGHQTCKNFKCKPLFWYCDSDVNDGSDNSDEGSCCPAQTRFCSNGKCLSK 540
QY 541 SQQNGKDDGSDGSDASCPKRVNVYTCIKHYRCLNGLCLSKGNPECDGKEDCSGSDSK 600
DB 541 SQQNGKDDGSDGSDASCPKRVNVYTCIKHYRCLNGLCLSKGNPECDGKEDCSGSDSK 600
QY 601 DCDGGLSFTROARVVGCTADDEGEMPMOYSLHALGCGHICGASLISPMNLVSAHACYID 660
DB 601 DCDGGLSFTROARVVGCTADDEGEMPMOYSLHALGCGHICGASLISPMNLVSAHACYID 660
QY 661 DRGFRYSDDPTQWTAFLGLHDOSQSRAPGVQERRLKRIISHPFNDFTFDYDIALLELEKP 720
DB 661 DRGFRYSDDPTQWTAFLGLHDOSQSRAPGVQERRLKRIISHPFNDFTFDYDIALLELEKP 720
QY 721 AEYSMWRIPICLPASHVFPAGKAIWYTGWHTQYGGTGALILQKGEIRVINTTCENL 780
DB 721 AEYSMWRIPICLPASHVFPAGKAIWYTGWHTQYGGTGALILQKGEIRVINTTCENL 780
QY 781 PQQITPRMCMCVGLSGVDSCQSDGSPSSVEADGRIFGAGVSWGDCAGRNKPGVYT 840
DB 781 PQQITPRMCMCVGLSGVDSCQSDGSPSSVEADGRIFGAGVSWGDCAGRNKPGVYT 840
QY 841 RLPLFRDWMIKENTGV 855
DB 841 RLPLFRDWMIKENTGV 855
```

```
RESULT 4
US-09-644-600-10
; Sequence 10, Application US/09644600
; Patent No. 6451500
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Tanimoto, Hirotoshi
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease
; FILE REFERENCE: D6064CIP/D
; CURRENT APPLICATION NUMBER: US/09/644,600
; CURRENT FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: 09/421,213
; PRIOR FILING DATE: 1999-10-20
; PRIOR APPLICATION NUMBER: 09/027,337
; PRIOR FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 10
; LENGTH: 902
```

TYPE: PRT
ORGANISM: Mus musculus
FEATURE:
OTHER INFORMATION: Epithin
US-09-644-600-10

Query Match 81.0%; Score 3793; DB 4; Length 902;
Best Local Similarity 81.1%; Pred. No. 4,86-269;
Matches 684; Conservative 71; Mismatches 88; Indels 0; Gaps 0;

```
QY 1 MGSRRARKGKGGGPKDGAAGLKYSRHRKVNGLSEGVFLPVNNVKKYKKGPRMVLAA 60
DB 1 MGSRRGRKAGGGQDPAAGLKYSRLNNMGFESEVFLPANNAKKYKKGPRMVLVA 60
QY 61 VLIGLLVLIGLIGFLVWHLQYRDVAVOKYKNGYRITNENFVDAYENSNSTEFVLSASKY 120
DB 61 VLSEFLLSLMAGLLVWHFHYRNVAVOKYKNGHRLITNEIFLDAYENSTSTEFVLSAQV 120
QY 121 KDAIKLLYSGVPLGPRYHKESAVTAFSSEGVIAAYWSEFSIPQHLVBEAERVMAEERVVM 180
DB 121 KEALIKLLYNEVPVLGPRYHKESAVTAFSSEGVIAAYWSEFSIPHLAEVDRAVMAVERVVT 180
QY 181 LPPRARSLSKFVVTSVVAFPTDSKTVOPTODNSCSFGLHARGVELMFTTPGPPSPYPA 240
DB 181 LPPRARSLSKFVVTSVVAFPTDPMLOPTODNSCSFALHARGAVALFTTPGPPNSPYPA 240
QY 241 HARCQWALRGDADSVLSLTFERSFDLASCDERGSDLVTVYNTLSPMEPHALVOLCGTYPSPS 300
DB 241 HARCQWALRGDADSVLSLTFERSFDVAPCDEHGSDLVTVYDLSLPMERHAYVRLCGTSPS 300
QY 301 YNLTFFSSQNVLLITLITNTERRHPRGFATFPQLPRMSSCGGLRKAQGTFFNSPYYPGHY 360
DB 301 YNLTFFLSSQNVFLVTLITNTERRHPRGFATFPQLPRMSSCGGLVLDPTQGTFFSSPYYPGHY 360
QY 361 PPNIDCTWNTNEVPNNQHVAVRKFYFLIEBPGVAGTCKDYVEINERKYGSEBSQFVYVS 420
DB 361 PPNINCTWNTKVPNNRNVKRFKLFYLVDPNVAVGSCTKDYVEINERKYGSEBSQFVYVS 420
QY 421 NSNKTIVRFHSDSYTDTGFLAEYLSYDSSDPGPGFTCTGRCIRKELRCDGMADCTDH 480
DB 421 NSSKITVYHFHSDHSYDTDTGFLAEYLSYDNDPCGMMCKTGRCIRKELRCDGMADCPDY 480
QY 481 SDELINCSGADGHQFTCKNKKFCKPLFWVCDSYNDGSDNSDEOGSCPAQOTFRCSNGKCLSK 540
DB 481 SDERYRCNATHQFTCKNKKFCKPLFWVCDSYNDGSDNSDEOGSCPAQGSFKCSNGKCLPQ 540
QY 541 SQQCKNGKDDGSDSDASCPKVVVVTCTKHTYRCLNGLCLSKNPECDGKEDSDGSDDEK 600
DB 541 SQQCKNGKDDGSDSDASCDSVNVVSCTKHTYRCQNGCLCLSKNPECDGKEDSDGSDDEK 600
QY 601 DCDGGLRSFTRQARVVGTDADGEMPMOYSLHALGQGHICGASLISPNMLVSAAHACYID 660
DB 601 NCCGGLRSFTRQARVVGTDADGEMPMOYSLHALGQGHICGASLISPMVLVSAAHCFOD 660
QY 661 DRGGRYSDPTQWTAFLGLHDOSQASAPGVQERRLKRIISHPFPNDFTFYDIALLELEKP 720
DB 661 DRKFYKSDYTMWTAFLGLHDOSKRSASAGVQELKRIITHPSPFNDFTFYDIALLELEKS 720
QY 721 AEXSSWVRPCLPASHVFPAGKAIWYTGWHTGYGTGALLQKGIIRVYNQTTCCNL 780
DB 721 VERSTVVRPCLPASHVFPAGKAIWYTGWHTGEGGTGALLQKGIIRVYNQTTCCDLM 780
QY 781 PQOITPRMNCVGLSGVDSCQDSSGGLPSSEADGRIFGAGVVSWMGDGAGRNKPPVYT 840
DB 781 PQOITPRMNCVGLSGVDSCQDSSGGLPSSEADGDMFQAGVVSWMGEGCAQRNKPVYT 840
QY 841 RLP 843
DB 841 RLP 843
```

RESULT 5
US-09-654-600A-10

Sequence 10, Application US/09654600A
Patent No. 6649741
GENERAL INFORMATION:
APPLICANT: O'Brien, Timothy J.
TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease
FILE REFERENCE: D6064CIP/D
CURRENT APPLICATION NUMBER: US/09/654,600A
PRIORITY FILING DATE: 2000-09-01
PRIORITY FILING DATE: 09/02/2001, 213
PRIORITY FILING DATE: 1999-10-20
NUMBER OF SEQ ID NOS: 98
SEQ ID NO 10
LENGTH: 902
TYPE: PRT
ORGANISM: Mus musculus
FEATURE:
OTHER INFORMATION: Epithin
US-09-654-600A-10

Query Match 81.0%; Score 3793; DB 4; Length 902;
Best Local Similarity 81.1%; Pred. No. 4,86-269;
Matches 684; Conservative 71; Mismatches 88; Indels 0; Gaps 0;

```
QY 1 MGSRRARKGKGGGPKDGAAGLKYSRHRKVNGLSEGVFLPVNNVKKYKKGPRMVLAA 60
DB 1 MGSRRGRKAGGGQDPAAGLKYSRLNNMGFESEVFLPANNAKKYKKGPRMVLVA 60
QY 61 VLIGLLVLIGLIGFLVWHLQYRDVAVOKYKNGYRITNENFVDAYENSNSTEFVLSASKY 120
DB 61 VLSEFLLSLMAGLLVWHFHYRNVAVOKYKNGHRLITNEIFLDAYENSTSTEFVLSAQV 120
QY 121 KDAIKLLYSGVPLGPRYHKESAVTAFSSEGVIAAYWSEFSIPQHLVBEAERVMAEERVVM 180
DB 121 KEALIKLLYNEVPVLGPRYHKESAVTAFSSEGVIAAYWSEFSIPHLAEVDRAVMAVERVVT 180
QY 181 LPPRARSLSKFVVTSVVAFPTDSKTVOPTODNSCSFGLHARGVELMFTTPGPPSPYPA 240
DB 181 LPPRARSLSKFVVTSVVAFPTDPMLOPTODNSCSFALHARGAVALFTTPGPPNSPYPA 240
QY 241 HARCQWALRGDADSVLSLTFERSFDLASCDERGSDLVTVYNTLSPMEPHALVOLCGTYPSPS 300
DB 241 HARCQWALRGDADSVLSLTFERSFDVAPCDEHGSDLVTVYDLSLPMERHAYVRLCGTSPS 300
QY 301 YNLTFFSSQNVLLITLITNTERRHPRGFATFPQLPRMSSCGGLRKAQGTFFNSPYYPGHY 360
DB 301 YNLTFFLSSQNVFLVTLITNTERRHPRGFATFPQLPRMSSCGGLVLDPTQGTFFSSPYYPGHY 360
QY 361 PPNIDCTWNTNEVPNNQHVAVRKFYFLIEBPGVAGTCKDYVEINERKYGSEBSQFVYVS 420
DB 361 PPNINCTWNTKVPNNRNVKRFKLFYLVDPNVAVGSCTKDYVEINERKYGSEBSQFVYVS 420
QY 421 NSNKTIVRFHSDSYTDTGFLAEYLSYDSSDPGPGFTCTGRCIRKELRCDGMADCTDH 480
DB 421 NSSKITVYHFHSDHSYDTDTGFLAEYLSYDNDPCGMMCKTGRCIRKELRCDGMADCPDY 480
QY 481 SDELINCSGADGHQFTCKNKKFCKPLFWVCDSYNDGSDNSDEOGSCPAQOTFRCSNGKCLSK 540
DB 481 SDERYRCNATHQFTCKNKKFCKPLFWVCDSYNDGSDNSDEOGSCPAQGSFKCSNGKCLPQ 540
QY 541 SQQCKNGKDDGSDSDASCPKVVVVTCTKHTYRCLNGLCLSKNPECDGKEDSDGSDDEK 600
DB 541 SQQCKNGKDDGSDSDASCDSVNVVSCTKHTYRCQNGCLCLSKNPECDGKEDSDGSDDEK 600
QY 601 DCDGGLRSFTRQARVVGTDADGEMPMOYSLHALGQGHICGASLISPNMLVSAAHACYID 660
DB 601 NCCGGLRSFTRQARVVGTDADGEMPMOYSLHALGQGHICGASLISPMVLVSAAHCFOD 660
QY 661 DRGGRYSDPTQWTAFLGLHDOSQASAPGVQERRLKRIISHPFPNDFTFYDIALLELEKP 720
DB 661 DRKFYKSDYTMWTAFLGLHDOSKRSASAGVQELKRIITHPSPFNDFTFYDIALLELEKS 720
```

```
Db 661 DKNFYSQTYTWTAFGLGLDQSKSASGVQELKRLITTHPSFNDFTFYDIALLELEKS 720
Qy 721 AEYSMWRLPCLPDASHVFPKAIWVTGHTQYGGTGAIILOKGEIRVINTTCENLL 780
Db 721 VEYSIVRPICLPATIHVPAGKAIWVTGHTQYGGTGAIILOKGEIRVINTTCEDLM 780
Qy 781 POQITPRMVCVGLSGVDSCQDSGGPLSSVEADGRIFGAGVSWGDCAGRNKPGVYT 840
Db 781 POQITPRMVCVGLSGVDSCQDSGGPLSSAEKDGRIFOAGVSWGDCAGRNKPGVYT 840
Qy 841 RLP 843
Db 841 RLP 843

RESULT 6
US-09-657-986B-2
; Sequence 2, Application US/09657986B
; Patent No. 6797504
; GENERAL INFORMATION:
; APPLICANT: Madison, Edwin L.
; APPLICANT: Semple, Joseph Edward
; APPLICANT: Coombs, Gary Samuel
; APPLICANT: Reiner, John Eugene
; APPLICANT: Ong, Edgar O.
; APPLICANT: Araldi, Gian Luca
; TITLE OF INVENTION: Inhibitors of Serine Protease Activity of Matricinase or
; FILE REFERENCE: Corvas 255/049
; CURRENT APPLICATION NUMBER: US/09/657, 986B
; CURRENT FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 2
; LENGTH: 241
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-09-657-986B-2

Query Match 27.9%; Score 1305; DB 4; Length 241;
Best Local Similarity 99.2%; Pred. No. 9.2e-88;
Matches 239; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 615 VVGSTDADEGEMPVQVSLHALGGGHCASLISPMWLVSAHCYDDGFRYSPTQWTA 674
Db 1 VVGSTDADEGEMPVQVSLHALGGGHCASLISPMWLVSAHCYDDGFRYSPTQWTA 60
Qy 675 FLGLHDSQSRAPGQVERLKRITISHPPFNFTFYDIALLELEKPAEYSSMWRLPCLPD 734
Db 61 FLGLHDSQSRAPGQVERLKRITISHPPFNFTFYDIALLELEKPAEYSSMWRLPCLPD 120
Qy 735 ASHVFPAGKAIWVTGHTQYGGTGAIILOKGEIRVINTTCENLLPOQITPRMVCVGL 794
Db 121 ASHVFPAGKAIWVTGHTQYGGTGAIILOKGEIRVINTTCENLLPOQITPRMVCVGL 180
Qy 795 SGGVDSGGGSGPLSSVEADGRIFGAGVSWGDCAGRNKPGVYTRPLPRDWIKENTG 854
Db 181 SGGVDSGGGSGPLSSVEADGRIFGAGVSWGDCAGRNKPGVYTRPLPRDWIKENTG 240
Qy 855 V 855
Db 241 V 241

RESULT 7
US-08-200-900A-2
; Sequence 2, Application US/08200900A
; Patent No. 565566
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
```

```
ADDRESSEE: Genetics Institute, Inc. - Legal Affairs
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: MA
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/200,900A
FILING DATE: 23-FEB-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Melnert, Maureen C.
REGISTRATION NUMBER: 31,544
REFERENCE/DOCKET NUMBER: GI 5201-FWC
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1170 X8574
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 798 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-200-900A-2

Query Match 15.0%; Score 700.5; DB 1; Length 798;
Best Local Similarity 32.6%; Pred. No. 7.2e-43;
Matches 181; Conservative 91; Mismatches 187; Indels 97; Gaps 24;

Qy 333 QLPKMSGCG--RLRKAQGTFSNPPYGHYPNIDCTWNIETVPNNQVKAKEFPYLLP 390
Db 298 ELP--TDGGPHDMEWNTFTSINFPNSVQNFCLMNLNAQKGKNIQLHFOGFDLENI 355
Qy 391 GVPAGTCPKDYVEI--NBEKTCGERSQV-----VTSNKNKTYTRHSQSTYDT 438
Db 356 A-----DVEIRDE--GDSLFLAVYTGPPVNDVFTNRMVTFETDMLAKQ 404
Qy 439 GFLAEVLS--YDSDDC--PQGFCTGTGRCIRKELRCGMDADCTDHSDELNC-----SC 488
Db 405 GFKNMFTTGGLGIPBECKEDNFQCKDGECLPLVNLCDGFPHCKDGSDEAHCVLFGITT 464
Qy 489 DAGH--OFTCKNKECKELFWVCDSVNDGDNSEDGSCCPAQTRFCSNGKCLSKSQOQNG 546
Db 465 DSSGLVQFRIS-----IMHV-----ACAEK-----WTQISDDVC-----QLIG 499
Qy 547 KDDCGDSDASG-----PKVNVVCTKHTYRCLNGLCLSKNPECDGKEDGSDSEK 600
Db 500 ---LGTNNSVPTRESTGGPYVNTAP-----NSSLITPQOQ-----LEDSLL 544
Qy 601 DCD---GGLSFTQA--RVVGTDADEGEMPVQVSLHALGGGHCASLISPMWLVSA 655
Db 545 QCNKSGCKKLVTQEVSPKIVGSSDSREGAIPWVALYFPDQ--QVCAISLSDRLVSA 603
Qy 656 HCYIDRGRYSPTQWTAFLGLHDSQSRAPGQVERLKRITISHPPFNFTFYDIAL 715
Db 604 HCVYG-----RMERSKKAVALGLMASNLISPOLIETLIDQIVLPHNKRKRKNDLMM 659
Qy 716 ELEKPAEYSSMWRLPCLPDASHVFPAGKAIWVTGHTQYGGTGAIILOKGEIRVINT 775
Db 660 HLEMKVYTDVQICLPEENQVPPPGRICSIAGMGALIVGSTADVLQEADEVLLSEK 719
Qy 776 CENLLPO--QITPRMVCVGLSGVDSCQDSGGPLSSVEADGRIFGAGVSWGDCAGRN 834
Db 720 CQOQMPENITENNVCAGYEAGVDSQDSGGPLMCOE--NNRMLLAGVTSFGIQCALPN 778
Qy 835 KPGVYTRPLPRDWIK 850
```

Db 779 RGVYARVPFTEMIQ 794

RESULT 8
US-08-794-042-2
Sequence 2, Application US/08794042
Patent No. 6746859
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc. - Legal Affairs
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: MA
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/794,042
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/200,900
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Meinert, Maureen C.
REGISTRATION NUMBER: 31,544
REFERENCE/DOCKET NUMBER: GI 5201-FWC
TELEPHONE: (617) 876-1170 X8574
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 798 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-794-042-2

Query Match 15.0%; Score 700.5; DB 4; Length 798;
Best Local Similarity 32.6%; Pred. No. 7.2e-43;
Matches 181; Conservative 91; Mismatches 187; Indels 97; Gaps 24;

Qy 333 QLRPMSSCGG--RLRKAQGTFSNPPYRGHYPPNIDCTWNIENVNNGHVKVRFKFFYLLEP 390
Db 298 ELP--TDCGGPHDLMBPNTFTSINFPNSYPNQAFCIMNLNAQKKNIOQLHFQEPDLENI 355
Qy 391 GVPAGTCKPKYVEI--NGEKYCGERSQFV-----VTSNKNITVRFHSDQSTYDT 438
Db 356 A-----DVVEIRGE---GDSLFPLAVYTGPGVNDVFSTNNMTVLFITDNLAKQ 404
Qy 439 GLPLAEYLS---YDSDPC--PGQFTCRGRCIRKELRCGDWADCTHSDLEINC-----SC 488
Db 405 GFRANFTTGYGLIGPECKEDNFQCKDGEICPLVNLCDGPFPHCKDSDEAHCVRLFNGT 464
Qy 489 DAGH--QFTCKNKFCKPLFWVCDNVNDGDNDSBOGSCPAQTFRCNSNGKCLSKSQOCNG 546
Db 465 DSSGLVQFRIQS-----IWHV-----ACAEIN-----WTTQISDDVC-----QLLG 499
Qy 547 KDCCGSDSDEASC-----PKVNVVCTKHTYRCLNGLCLSKGNPBCDGDGKSDGSDSEK 600
Db 500 ---LGTGNSSVPTFSFGGPGYVNLNTAP-----NGSLILTPSQOC-----LEDLILL 544
Qy 601 DCD---CGLSFTROA--RVVGTGTDADGEMPMQVSLHALGQGHICGASLISPWLVSA 655
Db 545 QCNYKSCGKKLVQEVSPKIVGSDSRREGAMPVVALYFPDQ-QVCGASLIVSRDMLVSA 603

Qy 656 HCYIDRGRFRYSDPTQMTAFGLHDSORSAPGVQERRLKRIISHPFNDFTEDYIAL 715
Db 604 HCYVG---RNMESPCKKAVLGLHMASNLTSPQIETRLDIQIYINHYNRRKANDIAM 659
Qy 716 ELKPAEYSMWAPICLPDASHVFPAGKAIWTTGKHGTQYGGALLIQGELRVINQTT 775
Db 660 HLEKKNVYTDYDIPCLPBNQVFPFGIRICSIAGMALIYQGSTAVLDEADVPLLSNEK 719
Qy 776 CENILPO-QTPPMWCVGLSGVSDSCQSDGSPGLSSVEADGRIFGAGVYMWGDCAGRN 834
Db 720 CQQMPERNITTEMVCAGYAGVSDSCQSDGSPGLMCP--NRMWLAGVTSFGYCALPN 778
Qy 835 KPGVYTRLPFLFRDWIK 850
Db 779 RGVYARVPFTEMIQ 794

RESULT 9
PCT-US94-00616-2
Sequence 2, Application PC/TUS9400616
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE
NUMBER OF SEQUENCES: 33
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/00616
FILING DATE:
CLASSIFICATION:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 798 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-00616-2

Query Match 15.0%; Score 700.5; DB 5; Length 798;
Best Local Similarity 32.6%; Pred. No. 7.2e-43;
Matches 181; Conservative 91; Mismatches 187; Indels 97; Gaps 24;

Qy 333 QLRPMSSCGG--RLRKAQGTFSNPPYRGHYPPNIDCTWNIENVNNGHVKVRFKFFYLLEP 390
Db 298 ELP--TDCGGPHDLMBPNTFTSINFPNSYPNQAFCIMNLNAQKKNIOQLHFQEPDLENI 355
Qy 391 GVPAGTCKPKYVEI--NGEKYCGERSQFV-----VTSNKNITVRFHSDQSTYDT 438
Db 356 A-----DVVEIRGE---GDSLFPLAVYTGPGVNDVFSTNNMTVLFITDNLAKQ 404
Qy 439 GLPLAEYLS---YDSDPC--PGQFTCRGRCIRKELRCGDWADCTHSDLEINC-----SC 488
Db 405 GFRANFTTGYGLIGPECKEDNFQCKDGEICPLVNLCDGPFPHCKDSDEAHCVRLFNGT 464
Qy 489 DAGH--QFTCKNKFCKPLFWVCDNVNDGDNDSBOGSCPAQTFRCNSNGKCLSKSQOCNG 546
Db 465 DSSGLVQFRIQS-----IWHV-----ACAEIN-----WTTQISDDVC-----QLLG 499
Qy 547 KDCCGSDSDEASC-----PKVNVVCTKHTYRCLNGLCLSKGNPBCDGDGKSDGSDSEK 600
Db 500 ---LGTGNSSVPTFSFGGPGYVNLNTAP-----NGSLILTPSQOC-----LEDLILL 544
Qy 601 DCD---CGLSFTROA--RVVGTGTDADGEMPMQVSLHALGQGHICGASLISPWLVSA 655
Db 545 QCNYKSCGKKLVQEVSPKIVGSDSRREGAMPVVALYFPDQ-QVCGASLIVSRDMLVSA 603
Qy 656 HCYIDRGRFRYSDPTQMTAFGLHDSORSAPGVQERRLKRIISHPFNDFTEDYIAL 715

Db 604 HCYVG----RMBESKMAVGLHMASSLTSPQIETRLIDQIVINPHYKKNNDIAMM 659
Qy 716 ELEKPAEYSSMVRPCLPDASHVFPAGKAIWYMGHTQYGTGALLIQKEIRVINGTT 775
Db 660 HLEKKNVNTDYIOPCLPEENQVPPPGICSIAGGALITYGSTDVDVIGADVPPLSNK 719
Qy 776 CENLLPQ-QITPRMVCVFLSGVDSQQGDSGGLSSVEADGRIFGAGVSMGDCAGRN 834
Db 720 CQOMPEINTENNVCAGYEAGVDSQQGDSGGLMCE-NNRMLLAGVTSFGYCALPN 778
Qy 835 KPGVYTRLPFRDWIK 850
Db 779 RGVYARVPRFTWIO 794

RESULT 10
US-09-734-675-4
; Sequence 4, Application US/09734675
; Patent No. 6365391
; GENERAL INFORMATION:
; APPLICANT: WEBSTER, Marion et al
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; FILE REFERENCE: CLO00862
; CURRENT APPLICATION NUMBER: US/09/734,675
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 407
; TYPE: PRT
; ORGANISM: Human
US-09-734-675-4

Query Match 12.4%; Score 583; DB 3; Length 407;
Best Local Similarity 45.9%; Pred. No. 1.2e-34;
Matches 119; Conservative 37; Mismatches 83; Indels 20; Gaps 6;

Qy 604 CGLR--SFTROANVGSTDADEGEWPMQVSLHLAGCHICGASLISPMVLVAHXYID 660
Db 162 CGTRRSKTLGSLRIVGTEVEGEPWQASLQMDG-SHRCGATLINATWLVSAHCT- 219
Qy 661 DRGRYSDPTQWTFELGHDQSQSAPOVERLKRITSHPFNDFTFYDIALLELEK 720
Db 220 ---TYKNPARKWTASFV---TTPSKMRKRLRIIVHEKXKHPSHDYDISLAELSSP 270
Qy 721 AEYSMVRPCLPDASHVFPAGKAIWYMGHTQYGTGALLIQKEIRVINGTTCEML 780
Db 271 VPTNNAHVRVCLPDASHVFPAGKAIWYMGHTQYGTGALLIQKEIRVINGTTCEML 328
Qy 781 PQ---QITPRMVCVFLSGVDSQQGDSGGLSSVEADGRIFGAGVSMGDCAGRNK 836
Db 329 PQAVNDATTPRMLCGSLGKGTDAQCQDGSGLVSSDARDIWLGIWMDGCAKPKP 388
Qy 837 GVTYTRLPFRDWIKENTGV 855
Db 389 GVTYTRVTRALRDMITSKGI 407

RESULT 11
US-09-685-166A-895
; Sequence 895, Application US/09685166A
; Patent No. 6630305
; GENERAL INFORMATION:
; APPLICANT: Xu, Jianshun
; APPLICANT: Dillon, Devin C.
; APPLICANT: Mitchem, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jia, Yugu
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.

; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedrick, Thomas S.
; APPLICANT: Carter, Patrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C21
; CURRENT APPLICATION NUMBER: US/09/685,166A
; CURRENT FILING DATE: 2000-10-10
; NUMBER OF SEQ ID NOS: 898
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 895
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-685-166A-895

Query Match 11.9%; Score 557.5; DB 4; Length 492;
Best Local Similarity 33.9%; Pred. No. 1.1e-32;
Matches 131; Conservative 57; Mismatches 129; Indels 69; Gaps 14;

Qy 522 GCSGPAQTRC-SNGKLSKSGQCNQDGDGSDSDESKRV--NTV----- 565
Db 110 GSKCSNGIECDSSGCTINPSMCDGVSHCPGGEDENRVRLYGPFLQMYSSQKSMH 169
Qy 566 -----TCT---KATYRCANG-----LCLSKGNPECDKEDCSD 595
Db 170 PVCDDPMNENYGRACACDMGKKNFYSSQGIIVDSSGTSFMKNTSAGNVDIYKLYHSD 229
Qy 596 GSDEK-----DCDGLRSFTRQARVGTADDEGEWPMQVSLHLAGCHICGASLIS 647
Db 230 ACSKAVSLRCLACGVNLNS-SRQSRIVGSESLPAPWPMQVSLH-VQVHVCGGSLIT 287
Qy 648 PNLVSAHACYIDRGRYSDPTQWTFELGHDQS-QSAPQVERLKRITSHPFND 706
Db 288 PEMVITVAHCEKPD---LNNPMHMTAFAGILRSFMYGAGYQ---VQKVISHPNYSK 340
Qy 707 TFDVIALLELEKPAEYSSMVRPCLPDASHVFPAGKAIWYMGHTQYGTGALLIQK 766
Db 341 TKANDILMLKQLTNDLVKPYCLPFGMMLQPELCHISGAGLEBKSTSEVLNAA 400
Qy 767 EIRVINGTTCEM-LPQOITPRMVCVFLSGVDSQQGDSGGLSSVEADGRIFG-AGV 823
Db 401 KVLIERORCNSRYVNDLITPAMICAGFLQGNVDSQQGDSGGL--VTSNNIMWLIGD 458
Qy 824 VSMGDCAGRNKPGVYTRLPFRDWI 849
Db 459 TSMGSCAKAYRPGVYGNVWVFTDWI 484

RESULT 12
US-09-879-792-14
; Sequence 14, Application US/09879792
; Patent No. 6734006
; GENERAL INFORMATION:
; APPLICANT: Xiao, Yonghong
; APPLICANT: Gedrich, Richard
; TITLE OF INVENTION: Regulation of Human Transmembrane Sertine
; TITLE OF INVENTION: Protease
; FILE REFERENCE: 02973.00035
; CURRENT APPLICATION NUMBER: US/09/879,792
; CURRENT FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/211,224
; PRIOR FILING DATE: 2000-06-13
; PRIOR APPLICATION NUMBER: US 60/283,353
; PRIOR FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: US 60/283,648
; PRIOR FILING DATE: 2001-04-16


```
; PRIOR APPLICATION NUMBER: PCT (Docket No. 6734006 LIO-81-WO)
; PRIOR FILING DATE: 2001-06-12
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-879-792-14

Query Match      11.9%; Score 557.5; DB 4; Length 492;
Best Local Similarity 33.9%; Pred. No. 1.1e-32;
Matches 131; Conservative 57; Mismatches 129; Indels 69; Gaps 14;

QY 522 GCSCPAQTRFC-SNGKCLSKSQOQNGKDDCGSDSDASCPKV---NVV----- 565
D 110 GSKCSNGIGIECDSSGTCINPNCMDGVSHCPGGEDENRVRLYGPNIILQVYSSQKSMH 169
QY 566 -----TCF---KHTYRCILNG-----LCLSKGNPECDGKEDCSD 595
D 170 PVCQDDMNENYGRAACRDMGKNNFYSSQGIYDSDGSTSFMKLNTSAGNVDIYKKLYHSD 229
QY 596 GSDEK-----DCDGLRSFTROARVVGSTADBEQWQVSLALGQCHICGASLIS 647
D 230 ACSSKAVVSLRCLACGVNLNS-SRQSRIVGSESLPGAMPQVSLH-VQNVHVCQGSIIIT 287
QY 648 PNVLYSAHCYIDRGFRYSDPTQWTAFLGLHDQS-ORSAFVQERRLKRIISHPFNDP 706
D 288 PEMIVTAHACVEXR-----LNNPMHMTAFAGILRQSFMYGAGYQ---VQKVISHPNYDSK 340
QY 707 TFDYDIALLELEKPAEYSSMWRIPLPDASHVPPAKAIWVTGNGHTQYGGTALILQKG 766
D 341 TKNNDIALMKLQKPLTFNDLVKPVCLPNPGMWLOPEQLCWSGGAETEKGKTSSEVLNAA 400
QY 767 EIRVINTQTCEN--LBPQITPRMVCYPLSGGVSDSCQSDSGPLSVYADGRIFG-AGV 823
D 401 KVLIIETQRNSRYVYDNLITPAMICAGFLQGNVSDSCQSDSGPL--VTSNNNIIMWLIGD 458
QY 824 VSWGDCGAGRNRKPGVYTRPLPLFRDWI 849
D 459 TSMGSCAKAKYRRPGVYGNVWVFETMI 484

RESULT 13
US-09-679-426-895
; Sequence 895, Application US/09679426
; Patent No. 6759515
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yugu
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Derrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C20
; CURRENT APPLICATION NUMBER: US/09/679,426
; NUMBER OF SEQ ID NOS: 895
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 895
```

```
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-679-426-895

Query Match      11.9%; Score 557.5; DB 4; Length 492;
Best Local Similarity 33.9%; Pred. No. 1.1e-32;
Matches 131; Conservative 57; Mismatches 129; Indels 69; Gaps 14;

QY 522 GCSCPAQTRFC-SNGKCLSKSQOQNGKDDCGSDSDASCPKV---NVV----- 565
D 110 GSKCSNGIGIECDSSGTCINPNCMDGVSHCPGGEDENRVRLYGPNIILQVYSSQKSMH 169
QY 566 -----TCF---KHTYRCILNG-----LCLSKGNPECDGKEDCSD 595
D 170 PVCQDDMNENYGRAACRDMGKNNFYSSQGIYDSDGSTSFMKLNTSAGNVDIYKKLYHSD 229
QY 596 GSDEK-----DCDGLRSFTROARVVGSTADBEQWQVSLALGQCHICGASLIS 647
D 230 ACSSKAVVSLRCLACGVNLNS-SRQSRIVGSESLPGAMPQVSLH-VQNVHVCQGSIIIT 287
QY 648 PNVLYSAHCYIDRGFRYSDPTQWTAFLGLHDQS-ORSAFVQERRLKRIISHPFNDP 706
D 288 PEMIVTAHACVEXR-----LNNPMHMTAFAGILRQSFMYGAGYQ---VQKVISHPNYDSK 340
QY 707 TFDYDIALLELEKPAEYSSMWRIPLPDASHVPPAKAIWVTGNGHTQYGGTALILQKG 766
D 341 TKNNDIALMKLQKPLTFNDLVKPVCLPNPGMWLOPEQLCWSGGAETEKGKTSSEVLNAA 400
QY 767 EIRVINTQTCEN--LBPQITPRMVCYPLSGGVSDSCQSDSGPLSVYADGRIFG-AGV 823
D 401 KVLIIETQRNSRYVYDNLITPAMICAGFLQGNVSDSCQSDSGPL--VTSNNNIIMWLIGD 458
QY 824 VSWGDCGAGRNRKPGVYTRPLPLFRDWI 849
D 459 TSMGSCAKAKYRRPGVYGNVWVFETMI 484

RESULT 14
US-09-342-749-2
; Sequence 2, Application US/09342749
; Patent No. 6166194
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Teng, David H.-F.
; APPLICANT: Myriad Genetics, Inc.
; TITLE OF INVENTION: TMRSS2 is a Tumor Suppressor
; FILE REFERENCE: 2318-202
; CURRENT APPLICATION NUMBER: US/09/342,749
; CURRENT FILING DATE: 1999-06-29
; EARLIER APPLICATION NUMBER: US 60/091,044
; EARLIER FILING DATE: 1998-06-29
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-342-749-2

Query Match      11.9%; Score 555.5; DB 3; Length 492;
Best Local Similarity 33.9%; Pred. No. 1.6e-32;
Matches 131; Conservative 56; Mismatches 130; Indels 69; Gaps 14;

QY 522 GCSCPAQTRFC-SNGKCLSKSQOQNGKDDCGSDSDASCPKV---NVV----- 565
D 110 GSKCSNGIGIECDSSGTCINPNCMDGVSHCPGGEDENRVRLYGPNIILQVYSSQKSMH 169
QY 566 -----TCF---KHTYRCILNG-----LCLSKGNPECDGKEDCSD 595
D 170 PVCQDDMNENYGRAACRDMGKNNFYSSQGIYDSDGSTSFMKLNTSAGNVDIYKKLYHSD 229
```

Query Match	11.9%	Score 555.5	DB 4	Length 492	
Best Local Similarity	33.9%	Pred. No. 1.6e-32			
Match 131	Conservative	56	Mismatches 130	Indels 69	Gaps 14
Qy	522	GCSCPAQCFRC-SNGKCLSKSQOQCKDKDCCGSDSEASCPKV---NVV-----	565		
Db	110	GSKCSNSGIECDSSSTCLINPSMWCGVSHCPGGEDENCVRLYGNFLIQVYSSQKSMH	169		
Qy	566	-----TCT---KHYRCLNG-----LCLSKGNPECDGKEDCSD	595		
Db	170	PVCDDMNENYGRACRDMGYKNNFYSSQGIYDSDGSTSFMKLNTSAGVNDIYKTLVHSD	229		
Qy	596	GSDEK-----DDCCLRFSTTRARVYGCYDADGEMPMQVSLHALGQHILCGASILS	647		
Db	230	ACSSAAVYSLRCLACGVNLNS-SRQSRIRYVGSGSALPAMPQVSLH-VQNVHVCQSIT	287		
Qy	648	PNMLVSAACHYIDRGFRYSDPTQWTAFAGLIDQS-QRSAPGVQERRLRILSHPFDF	706		
Db	288	PEWITAAHCHVEP---LNNPMHTTAFAGILRGFMFGAQY---VEKVIASHNNYDSK	340		
Qy	707	TFDDYDIALLEKPAEYSSMWAPICLPDASHVFPAGKAIWTTGWHGTGYGTGAILQKG	766		
Db	341	TKNNDIALMKLOKPLTFNDLVKPVCLPNCMMLOEQLCWMGATEBKGTSEVLNAA	400		
Qy	767	EIRVYNQTTCCN-LLPQOITPRMNCVGTLSGVNVSQQDSSGGRPLSSVEADRIIG-AGV	823		
Db	401	KYLLLETQRKNSRYVDNLITPAMIQAGLQGVNVSQQDSSGGRPL-VYSKNNIWMILGD	458		

```

US-08-944-483-65
; Sequence 65, Application US/08944483
; Patent No. 6232456
; GENERAL INFORMATION:
; APPLICANT: COHEN, MAURICE
; APPLICANT: COLPITS, TRACEY L.
; APPLICANT: FRIEDMAN, PAULA N.
; APPLICANT: GRAMADOS, EDWARD N.
; APPLICANT: KLASS, MICHAEL R.
; APPLICANT: RUSSELL, JOHN C.
; APPLICANT: STEWART, KENT D.
; APPLICANT: STROUPE, STEVEN D.
; TITLE OF INVENTION: NOVEL SERINE PROTEASE REAGENTS
; TITLE OF INVENTION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES
; TITLE OF INVENTION: OF THE PROSTATE
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Road
; CITY: Abbott Park
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/944,483
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Becker, Cheryl L.
; REGISTRATION NUMBER: 35,441
; REFERENCE/DOCKET NUMBER: 6183.US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 847/935-1729
; TELEFAX: 847/938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 235 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 6232456e
US-08-944-483-65

Query Match      11.6%; Score 543.5; DB 3; Length 235;
Best Local Similarity 44.7%; Pred. No. 4.6e-32;
Matches 106; Conservative 44; Mismatches 80; Indels 7; Gaps 4;

QY      615 VGGSTADDEGEMPNVSLHLAGGHI CGASLISPMWLVAACHYIDDRGRYSDDPTOWTA 674
DB      1 IVGSNNKEAAMPVVGLI-YGRLICGASLVSSDMLVAACVTG---RULEFSKMTA 55
QY      675 FLAGLDQSASABGVDERRLKRIISHPFNDFTFYDIALLLELRPAEYSSMWRCICLPD 734
DB      56 ILLGHMSNLTSGQTYPRLDIVINPHYNRRKNDIMMHLFPKVNTDYIQICLPE 115
QY      735 ASHWEPAGKALVWTGNGHTOYGGTGALLIOKGSEIRPINOTGTENLLPQ-QITPPRMCTGF 793
       : |||::: | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |

```

Db 116 ENQVFPGRNCISLAWGTIVVYQGTIANILQADVPLLSNRCCQOMBEYNITENMCAGY 175

QY 794 LSGGVDSGCGDSGGPPLSSVEADGRIFGAGVYSWGDGAGRNKPGVYTRLPPLFRDWIK 850

Db 176 EAGGVDSGCGDSGGPPLMCQE-NNRWFLAGVTSFGYQCALPVRPGVYARVPFTWIO 231

RESULT 17

US-08-807-151-3

; Sequence 3, Application US/08807151

; Patent No. 6043033

; GENERAL INFORMATION:

; APPLICANT: Bandman, Olga

; APPLICANT: Lal, Preeti

; TITLE OF INVENTION: NOVEL HUMAN PROSTATE-ASSOCIATED

; TITLE OF INVENTION: PROTEASE

; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Drive

; CITY: Palo Alto

; STATE: CA

; COUNTRY: US

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/807.151

; FILING DATE: Filed Herewith

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Billings, Lucy J.

; REGISTRATION NUMBER: 36,749

; REFERENCE/DOCKET NUMBER: PF-0227 US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-855-0555

; TELEFAX: 415-845-4166

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 235 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; LIBRARY: GenBank

; CLONE: 416132

US-08-807-151-3

Query Match 11.5%; Score 540.5; DB 3; Length 235;

Best Local Similarity 43.5%; Pred. No. 7.6e-32;

Matches 103; Conservative 50; Mismatches 77; Indels 7; Gaps 4;

QY 615 VVGSTDADEGEWQVSLHALGQHICGASLISPNMLVSAHCVYIDDRGRYSDPQWTA 674

Db 1 IVGGSDBREGAMWVVALYFDDQ-QVCGASLVSRDWLVSAHCVYG---RNNEPSKMK 55

QY 675 FLGHDSQSAPQVQERRKRIISHPFNDFTFDYDIALLEKPAEYSSMVRPICLPD 734

Db 56 VLGHMASNLTSQIETRLIDQIVINPHYKRRKNDIAMHLEMKVNTDYIOPICLPE 115

QY 735 ASHVFPAGKAIWYTGHTGYGCTGALLIQKEIRVINQTTCEMLPQ-QITPRMVCVF 793

Db 116 ENQVFPGRNCISLAWGTIVVYQGTIANILQADVPLLSNRCCQOMBEYNITENMCAGY 175

QY 794 LSGGVDSGCGDSGGPPLSSVEADGRIFGAGVYSWGDGAGRNKPGVYTRLPPLFRDWIK 850

Db 176 EAGGVDSGCGDSGGPPLMCQE-NNRWFLAGVTSFGYQCALPVRPGVYARVPFTWIO 231

RESULT 18

US-09-478-957-3

; Sequence 3, Application US/09478957

; Patent No. 6350448

; GENERAL INFORMATION:

; APPLICANT: Bandman, Olga

; APPLICANT: Lal, Preeti

; TITLE OF INVENTION: NOVEL HUMAN PROSTATE-ASSOCIATED

; TITLE OF INVENTION: PROTEASE

; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Drive

; CITY: Palo Alto

; STATE: CA

; COUNTRY: US

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/478,957

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/807,151

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Billings, Lucy J.

; REGISTRATION NUMBER: 36,749

; REFERENCE/DOCKET NUMBER: PF-0227 US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-855-0555

; TELEFAX: 415-845-4166

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 235 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; LIBRARY: GenBank

; CLONE: 416132

US-09-478-957-3

Query Match 11.5%; Score 540.5; DB 3; Length 235;

Best Local Similarity 43.5%; Pred. No. 7.6e-32;

Matches 103; Conservative 50; Mismatches 77; Indels 7; Gaps 4;

QY 615 VVGSTDADEGEWQVSLHALGQHICGASLISPNMLVSAHCVYIDDRGRYSDPQWTA 674

Db 1 IVGGSDBREGAMWVVALYFDDQ-QVCGASLVSRDWLVSAHCVYG---RNNEPSKMK 55

QY 675 FLGHDSQSAPQVQERRKRIISHPFNDFTFDYDIALLEKPAEYSSMVRPICLPD 734

Db 56 VLGHMASNLTSQIETRLIDQIVINPHYKRRKNDIAMHLEMKVNTDYIOPICLPE 115

QY 735 ASHVFPAGKAIWYTGHTGYGCTGALLIQKEIRVINQTTCEMLPQ-QITPRMVCVF 793

Db 116 ENQVFPGRNCISLAWGTIVVYQGTIANILQADVPLLSNRCCQOMBEYNITENMCAGY 175

QY 794 LSGGVDSGCGDSGGPPLSSVEADGRIFGAGVYSWGDGAGRNKPGVYTRLPPLFRDWIK 850

Db 176 EAGGVDSGCGDSGGPPLMCQE-NNRWFLAGVTSFGYQCALPVRPGVYARVPFTWIO 231

RESULT 19

US-09-518-046-2

; Sequence 2, Application US/09518046

QY 56 CTGATATCTNLGTCLSGNGNECDGKECCSPGSEKDCGSLSPFROAVVGTADAGEW 626

Db 122 XRSSXKXCSGXVSL-----QCS----XQCGVRNLAAKMTSIVGGXASXGXW 168

QY 627 PMOVSJLHAGCGHITGCASLISPMLVSAKCYTIDDRGRYSDP--TQMTAFGLHDGQRS 685

Db 169 PMWSTLQKXXGVHLCCGSLIXPWWLTAHCVXG---RXKKPLXGMVFGIITJSLHX 224

QY 686 APGVQERRLRKRIISHPPNDFTF-----DYDIALLELEKPAEYSMWPRJCLPDASHVF 739

Db 225 XPLAAXRXRYEKIITHENYXXXXXXXSKXNDNDIALMKSKPLTFXDYIQPCLENPQXL 284

QY 740 PAKKAIWVTGNGHT--QYGGTGAALLQKEIRIVNOTGEN--LIPQOITPRMVCVGLSG 766

Db 285 XPEITTCMIXMGATETEXEKSTSPVLOEAAVPLIDMKNCNSXYVDNKTTPRMICAGYLEG 344

QY 797 GVDSCGGDSGGPL---SSVEADGRIFGAGVSWGGGCGAGR--NKGVRYRLPLPFDWI 849

Db 345 GVDSCGGDSGGPLVCEXXXXXQNNRWWLKGXTISWGACAKANRPPVYTXKTXPLKWI 401

RESULT 22
US-08-807-151-1

```

GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Lal, Preeti
TITLE OF INVENTION: NOVEL HUMAN PROSTATE-ASSOCIATED
TITLE OF INVENTION: PROTEASE
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSER: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/807,151
FILING DATE: Filed Herewith
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0227 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 283 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: SCORNO701
CLONE: 556016

```

Query Match	10.9%;	Score 509.5;	DB 3;	Length 283;
Best Local Similarity	39.5%;	Pred. No. 1.8e-29;		
Matches 11;	Conservative 47;	Mismatches 100;	Indels 23;	Gaps 9

```

Oy      581 SKNPEDCKEDKCCSGSGSEK-----DCCCGRSFTRQARVVGCTADAGEPMQVSL 6322
Db      6  SANNVDIYKKLHNSDACSKAAVSLRACIAGVNLNS--SRGSRITGGSALPGAMPQVSL 64

Oy      633 HALGCHITGASLISBNVLVLSAAHCTIDDRGFRYSDPQTORTAFGLGHDG--QSSAEGVGE 651
Db      65  H-VQNHHVCGSGIITPEMVLTPAAHVEKPE---LNNPHMHTAFEGILLRQSPFMFYGAGYQ- 118

Oy      692 RLKRLTIISHPFNDFPFDDIALLELEKPAEYSSMVAEPICLDPAASHVFPAGKAITWYGMG 751
Db      119 --VEKTIISHPNDSKTKNDIALMTLQKPLFNPDLYKPVCLPKNMGMLQGEOLCMTSGMG 176

Oy      752 HTQYGGTGAIIQKEIRIVNOTTCEN--LLPQOITTPMNCVGFSLSGVDSCGGDSGGPL 809
Db      177 ATEEEKKTSEVILNAAKVLIETFRQNSRKYVDNLTIPAMICAGPLQGVNDSCOGDSGGXL 236

Oy      810 SSVEADGRIFG--AGVAVMGDCGAGNKRGVYTRLPPLPDMT 849
Db      237 --VTSKNINIMWLIGDTSMGSCGCAKYRGGVYGNVWFIDMT 275

```

RESULT 23
US-09-478-957-1
; Sequence 1, Application US/09478957

```

1  APPLICANT: Bandman, Olga
2  APPLICANT: Lal, Preeti
3  TITLE OF INVENTION: NOVEL HUMAN PROSTATE-ASSOCIATED
4  TITLE OF INVENTION: PROTEASE
5  NUMBER OF SEQUENCES: 5
6  CORRESPONDENCE ADDRESS:
7  ADDRESSEE: Incyte Pharmaceuticals, Inc.
8  STREET: 3174 Porter Drive
9  CITY: Palo Alto
10 STATE: CA
11 COUNTRY: US
12 ZIP: 94304
13 COMPUTER READABLE FORM:
14 MEDIUM TYPE: Diskette
15 COMPUTER: IBM Compatible
16 OPERATING SYSTEM: DOS
17 SOFTWARE: FastSeq Version 2.0
18 CURRENT APPLICATION DATA:
19 APPLICATION NUMBER: US/09/478,957
20 FILING DATE:
21 CLASSIFICATION:
22 PRIOR APPLICATION DATA:
23 APPLICATION NUMBER: 08/807,151
24 FILING DATE:
25 ATTORNEY/AGENT INFORMATION:
26 NAME: Billings, Lucy J.
27 REGISTRATION NUMBER: 36,749
28 REFERENCE/DOCKET NUMBER: PF-0227 US
29 TELECOMMUNICATION INFORMATION:
30 TELEPHONE: 415-855-0555
31 TELEFAX: 415-845-4166
32 INFORMATION FOR SEQ ID NO: 1:
33 SEQUENCE CHARACTERISTICS:
34 LENGTH: 283 amino acids
35 TYPE: amino acid
36 STRANDEDNESS: single
37 TOPOLOGY: linear
38 IMMEDIATE SOURCE:
39 LIBRARY: SCORNO701
40 CLONE: 556016
41 US-09-478-957-1

```

Query Match	10.9%	Score 509.5;	DB 3;	Length 283;
Best Local Similarity	39.5%	Pred. No. 1.8e-29;		
Matches 111;	Conservative 47;	Mismatches 100;	Indels 23;	Gaps 9
581	SKGNPECDGDCGDCSDSEK-----DCDCGARSFTQRARVVGCTADAGSEWPMQVSL			
632				

Db 6 SAGVNDIYKLYHSDAGSSKAVLSRCLACGVNLNS-SRQSRIVGSEBALGAMPVQSL 64
Qy 633 HALGQGHICGASLISPMNLVSAHCYIDDRGFRYSPTQMTAFGLHDQS-QRSAPGVQE 691
Db 65 H-VQNVHVCSSGSIITPEWITVAACHVEKP-----LNNPMHMTAFAGILQSFMFYGAGYQ- 118
Qy 692 RLRLRIISHPPFNFTPDYDIALLEKPAVSSMVRPICPDASHVPAKAIWTVQWG 751
Db 119 --VERKIVSHPNYDSKTKNNDIALMKLQKPLTFNDLVKVCJPNPQMMLOPEOLCMISGMG 176
Qy 752 HTQYGTGALILQKEIRIVNOTTCEN--LLPQOITPRMVCVFLSGVSDSCQDSGGPL 809
Db 177 ATEKRGKTSVTLNAKVLITETQKNSRYVDNLITPAMICAGFLQGVSDSCQDSGGXL 236
Qy 810 SSVSEADGRIFG-AGVWSGDCAGRNKPGVYTRLLPLFRDWI 849
Db 237 --VTSKNIMWMLIGDTSWGSACAKAYRPGVYGNVWVFTDWI 275

RESULT 24
US-10-177-661-2
; Sequence 2, Application US/10177661
; Patent No. 6794173
; GENERAL INFORMATION:
; APPLICANT: Anderson, Dirk M.
; TITLE OF INVENTION: DENDRITIC CELL TRANSMEMBRANE SERINE PROTEASE
; FILE REFERENCE: 3256-A
; CURRENT APPLICATION NUMBER: US/10/177,661
; CURRENT FILING DATE: 2002-06-20
; PRIOR APPLICATION NUMBER: US 60/299,606
; PRIOR FILING DATE: 2001-06-20
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 2
; LENGTH: 477
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-177-661-2

Query Match 10.9%; Score 509.5; DB 4; Length 477;
Best Local Similarity 32.3%; Pred. No. 3.5e-29;
Matches 134; Conservative 50; Mismatches 133; Indels 93; Gaps 17;

Qy 464 CIRKELRCGWDCTDHSDELNC-----SCDAGHQTCKNFKCKPLFW--VCD 509
Db 114 CPKHAVRCDGVVDCIKLSDELGCVRFMDKSLIKIYSGSSHQ-----WLPICS 161
Qy 510 SVNDGNSDDEQSCGCAQFRCSNGKCLSKSQCCNGKDDCGSDSDEASCPKVVVYCTK 569
Db 162 S--WNDSYSSEKTC-----QQLGESAHRTTEVHRDPANFSILR 200
Qy 570 HTYRCLNGLCLSKGNPECDGKE---DCSDGSEKDCDCGLRSFTRQARVVGTDADGE 625
Db 201 YNSTIQESLHRS---ECPQRYISLQCS-----HCGLRAMT--GRIVGALASDSK 246
Qy 626 WPMQVSLHALGQGHICGASLISPMNLVSAHCYIDDR-----GFRYSPTQMTAFGLG--- 677
Db 247 WPMQVSLH-FGTTHICGCTLIDAQWVLTAAHCFVTRKYLEG-----WKVYAGTSN 297
Qy 678 LHDQSQSAAGVQERRLKRIISHPPFNFTPDYDIALLEKPAVSSMVRPICLPDASH 737
Db 298 LHQPEAS-----IAEIIINSNYTDEEDVDIALMRLSKPLTSLAHHPACLPWNGQ 350
Qy 738 VFPAKAIWTVGWT-OYGTGALILQKEIRIVNOTTCENLL--PQOITPRMVCVFL 794
Db 351 TFSINETCWTITGFKTRETDDKTSPLREAVQVNLIDPKKNDYLVYDSYILPRMCAGDL 410
Qy 795 SGGVSDSCQDSGGPLSSVEADGRIFGAGVWSGDCAGRNKPGVYTRLLPLFRDWI 849
Db 411 RGRDSCQDSGGPL--VCEQNNRWYLAGVTSWGTGCGGRNKPGRVYTKVTEVLPMI 464

RESULT 25
US-09-879-792-12
; Sequence 12, Application US/09879792
; Patent No. 6734006
; GENERAL INFORMATION:
; APPLICANT: Xiao, Yonghong
; TITLE OF INVENTION: Regulation of Human Transmembrane Serine
; FILE REFERENCE: 02973.00035
; CURRENT APPLICATION NUMBER: US/09/879,792
; CURRENT FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/211,224
; PRIOR FILING DATE: 2000-06-13
; PRIOR APPLICATION NUMBER: US 60/283,353
; PRIOR FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: US 60/283,648
; PRIOR FILING DATE: 2001-04-16
; PRIOR APPLICATION NUMBER: PCT
; SEQ ID NO 12
; LENGTH: 562
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-879-792-12 (Docket No. 6734006 LIO-81-WO)

Query Match 10.9%; Score 509.5; DB 4; Length 562;
Best Local Similarity 32.3%; Pred. No. 4.3e-29;
Matches 134; Conservative 50; Mismatches 138; Indels 93; Gaps 17;

Qy 464 CIRKELRCGWDCTDHSDELNC-----SCDAGHQTCKNFKCKPLFW--VCD 509
Db 199 CPKHAVRCDGVVDCIKLSDELGCVRFMDKSLIKIYSGSSHQ-----WLPICS 246
Qy 510 SVNDGNSDDEQSCGCAQFRCSNGKCLSKSQCCNGKDDCGSDSDEASCPKVVVYCTK 569
Db 247 S--WNDSYSSEKTC-----QQLGESAHRTTEVHRDPANFSILR 285
Qy 570 HTYRCLNGLCLSKGNPECDGKE---DCSDGSEKDCDCGLRSFTRQARVVGTDADGE 625
Db 286 YNSTIQESLHRS---ECPQRYISLQCS-----HCGLRAMT--GRIVGALASDSK 331
Qy 626 WPMQVSLHALGQGHICGASLISPMNLVSAHCYIDDR-----GFRYSPTQMTAFGLG--- 677
Db 332 WPMQVSLH-FGTTHICGCTLIDAQWVLTAAHCFVTRKYLEG-----WKVYAGTSN 382
Qy 678 LHDQSQSAAGVQERRLKRIISHPPFNFTPDYDIALLEKPAVSSMVRPICLPDASH 737
Db 383 LHQPEAS-----IAEIIINSNYTDEEDVDIALMRLSKPLTSLAHHPACLPWNGQ 435
Qy 738 VFPAKAIWTVGWT-OYGTGALILQKEIRIVNOTTCENLL--PQOITPRMVCVFL 794
Db 436 TFSINETCWTITGFKTRETDDKTSPLREAVQVNLIDPKKNDYLVYDSYILPRMCAGDL 495
Qy 795 SGGVSDSCQDSGGPLSSVEADGRIFGAGVWSGDCAGRNKPGVYTRLLPLFRDWI 849
Db 496 RGRDSCQDSGGPL--VCEQNNRWYLAGVTSWGTGCGGRNKPGRVYTKVTEVLPMI 549

RESULT 26
US-08-508-448C-25
; Sequence 25, Application US/08508448C
; Patent No. 5804410
; GENERAL INFORMATION:
; APPLICANT: Kazuyoshi YAMAKA et al.
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCE ENCODING
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack

QY 753 TQVGTGALLIKQGEIRIVINOTTCELLPLQO-----ITPRMVCYGFSLGGVDSQQSGSG 808
 Db 314 QEVAGHTVPELRQGGVRIISNDVCN--APHSTNGAILSTGMLCAGVPGQGVDAQCQSSGSG 371
 QY 809 LSSVPEADQR--IFGAGVVSWMQPCGCAKRNKPGYTRRLPLEFRDMIKENTGV 855
 Db 372 L--VQEDSRRLMFIIVIGIVSMWQCGPLPDPKPGYTRTRATALLDMIRQDTGI 418

```

US-09-854-133-82
RESULT 29
/ Sequence 82, Application US/09854133
/ Patent No. 6759508
/ GENERAL INFORMATION:
/ APPLICANT: Lodges, Michael J.
/ APPLICANT: Mohmamed, Raodoh
/ APPLICANT: Henderson, Robert A.
/ APPLICANT: Benson, Darin R.
/ APPLICANT: Secist, Heather
/ TITLE OF INVENTION: COMPOSITIONS AND METH
/ TITLE OF INVENTION: THE THERAPY AND DIAG
/ FILE REFERENCE: 210121.475C10
/ CURRENT APPLICATION NUMBER: US/09/854,133
/ CURRENT FILING DATE: 2001-05-11
/ NUMBER OF SEQ ID NOS: 735
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 82
/ LENGTH: 418
/ TYPE: prt
/ ORGANISM: Homo sapien
US-09-854-133-82

```

Query Match	10.9%	Score 508.5;	DB 4;	Length 418;
Best Local Similarity	37.4%;	Pred. No. 3.5e-29;		
Matches 108; Conservative	52;	Mismatches 100;	Indels 29;	Gaps 10

[illegible]

RESULT 30 113-83
US-09-854-133-83
Sequence 83, Application US/09854133
Patent No. 6759508
GENERAL INFORMATION:
APPLICANT: Lodes, Michael J.
APPLICANT: Mohamath, Radooh
APPLICANT: Henderson, Robert A.
APPLICANT: Benson, Darin R.
APPLICANT: Secrist, Heather
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
TITILE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
FILE REFERENCE: 210121.475C10
CURRENT APPLICATION NUMBER: US/09/854,133
CURRENT FILING DATE: 2001-05-11

```

; NUMBER OF SEQ ID NOS: 735
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO: 83
; LENGTH: 418
; TYPE: FRT
; ORGANISM: Homo sapien
US-09-854-133-83

```

Query Match	10.9%;	Score 508.5;	DB 4;	Length 418;
Best Local Similarity	37.4%;	Pred. No. 3.5e-29;		
Matches 108; Conservative	52;	Mismatches 100;	Indels 29;	Gaps 10;

Qy	581	SKNPEDCKEDKCSGSGSE-----KCCDGLSFT--RQARVVGVDADAGEKMPQVSL	632
Db	145	NSNMLEINSTETLSLTQDAANWVLINCGAGPDLITISEGRILIGTEABEGSGMPQVSL	204
Qy	633	HALGQGHITGASLISENVLVSAACFYIDDRGFRYSDPQTWATFLGLHDGORSAPYQER	692
Db	205	R--LNNAHRCGGSLINMMWILTLAAHCFRBN-----SNPDWATATGI-----STTFPKLR--M	253
Qy	693	RLKRIISHPFNDFPDVDIALLELEKDAEYSMWPRCLCPDPAHVFEPAGKAIWVTMGH	752
Db	254	RVNRIILHNHYSAHTENDIALVRLENSVTYTKDISHCSLPAAQNIPRGSTAYVTMGCA	313
Qy	753	TOYGTGALLQKGBIRVINQTTCEMLLPQO-----ITPRMVCVFLSGVDSCGGDSGGP	808
Db	314	QBYAGHTVPELRKQGVRIISDVCN--APHSYNGLIISGMCLAGVPQGVADACGGDSGGP	371
Qy	809	LSSVEADGR--IFGAGVYSWKGDCGARRKPPVYTRPLFRPMIKENTGV	855
Db	372	L--VQDSRRLLPFIQVIVSWDQCLPKRPQVYTRVATVADMIKQQGTGI	418

Search completed: November 29, 2004, 08:34:25
Job time : 35.2432 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 29, 2004, 08:14:53 ; Search time 122.858 Seconds
(without alignments)
2496.496 Million cell updates/sec

Title: us-09-936-333-27

Perfect score: 4683
Sequence: 1 MSGDRAKRGCGGPKRFGAGL.....PGVYTRLPFRDMIKENTGV 855

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729239 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 100 summaries

Database :

1: Geneseq_23Sep04:*
2: geneseqp1980s:*
3: geneseqp1990s:*
4: geneseqp2000s:*
5: geneseqp2001s:*
6: geneseqp2002s:*
7: geneseqp2003as:*
8: geneseqp2003bs:*
9: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4659	99.5	855	3 AAB19552	Aab19552 Human mat
2	4659	99.5	855	4 AAB35465	Aab35465 Human mem
3	4659	99.5	855	5 ADI16817	Adi16817 Human NOV
4	4659	99.5	855	5 ADI16883	Adi16883 Human NOV
5	4659	99.5	855	5 ADI16876	Adi16876 Human NOV
6	4659	99.5	855	7 ADN13967	Adn13967 Cancer/an
7	4659	99.5	855	8 ADN04754	Adn04754 Antiproti
8	4655	99.4	855	5 ADI16884	Adi16884 Human NOV
9	4655	99.4	855	5 ADI16818	Adi16818 Human NOV
10	4653	99.4	855	2 AAY06671	Aay06671 Tumour an
11	4653	99.4	855	4 AAB98500	Aab98500 Human TAD
12	4653	99.4	855	4 AAE06930	Aae06930 Human mem
13	4653	99.4	855	5 AAO23929	Aao23929 Type II t
14	4653	99.4	855	5 ADI16816	Adi16816 Human NOV
15	4653	99.4	855	5 ADI16882	Adi16882 Human NOV
16	4653	99.4	855	5 ADI16875	Adi16875 Human NOV
17	4653	99.4	855	6 AAB56619	Aab56619 Human mem
18	4653	99.4	855	6 AAO30146	Aao30146 Human mem
19	4653	99.4	855	6 AAE29820	Aae29820 Human mem
20	4653	99.4	855	6 AAE29791	Aae29791 Human mem
21	4653	99.4	855	6 AEP72376	Aep72376 Transmemb
22	4653	99.4	855	7 ADB97551	Adb97551 Human MTS
23	4653	99.4	855	7 ADI10371	Adi10371 Human cel
24	4653	99.4	855	8 ADG65326	Adg65326 Human MTS
25	4653	99.4	855	8 ADI28861	Adi28861 Human mat

26	4653	99.4	855	8 ADJ46895	Adj46895 Human tra
27	4625	98.8	853	8 ADO55145	Ado55145 Protein #
28	4614	98.5	851	4 AAM25628	Aam25628 Human pro
29	4614	98.5	851	4 ABB11428	Abb11428 Human mem
30	4302	91.9	932	4 ABG21442	Abg21442 Novel hum
31	4167	89.0	762	3 AAY90284	Aay90284 Human pep
32	4158.5	88.8	782	5 ABG96427	Abg96427 Human ova
33	4092	87.4	757	5 ADI16508	Adi16508 Human NOV
34	4057	86.6	757	8 ADN42162	Adn42162 Human nov
35	3884	82.9	855	5 AAE23083	Aae23083 Epithin p
36	3884	82.9	855	5 ADI16819	Adi16819 Murine NO
37	3864	82.6	855	5 ADI16877	Adi16877 Murine NO
38	3866	82.6	855	5 ADI16820	Adi16820 Rat NOVX
39	3866	82.6	855	5 ADI16878	Adi16878 Rat NOVX
40	3866	82.6	855	4 AAB98507	Aab98507 Murine ep
41	3793	81.0	902	5 AAB98517	Aab98517 Mouse epi
42	3793	81.0	902	5 AAB98517	Aab98517 Murine ty
43	3773	80.6	633	3 AAB19551	Aab19551 Human mat
44	3773	80.6	633	3 AAB19551	Aab19551 Human mat
45	2986	63.8	620	3 AAB37748	Aab37748 Human can
46	2660	56.8	845	5 ADI16879	Adi16879 African c
47	1343	28.7	362	4 ABG21441	Abg21441 Novel hum
48	1305	27.9	241	4 AAE06936	Aae06936 Human mem
49	1305	27.9	241	5 AAO22930	Aao22930 Type II t
50	1305	27.9	241	5 AAE22837	Aae22837 Human mat
51	1305	27.9	241	6 AAB56620	Aab56620 Human MTS
52	1305	27.9	241	6 AAO30147	Aao30147 Human mem
53	1305	27.9	241	6 AAE29821	Aae29821 Human mem
54	1305	27.9	241	6 AAE29792	Aae29792 Human mem
55	1305	27.9	241	6 AEP72377	Aep72377 Transmemb
56	1305	27.9	241	7 ADA25176	Ada25176 Human MTS
57	1305	27.9	241	7 ADB97553	Adb97553 Human MTS
58	1305	27.9	241	8 ADI28862	Adi28862 Human mat
59	1301	27.8	241	8 AAM22987	Aam22987 Human ser
60	1117.5	23.9	795	5 AEP43952	Aep43952 Human PRO
61	1114.5	23.8	799	5 ADI16874	Adi16874 Murine NO
62	1114.5	23.8	799	5 ADI16880	Adi16880 Murine NO
63	1112	23.7	802	2 AAY41710	Aay41710 Human PRO
64	1112	23.7	802	3 AAB44266	Aab44266 Human PRO
65	1112	23.7	802	3 AAB24052	Aab24052 Human PRO
66	1112	23.7	802	5 AAB27555	Aab27555 Amlno aci
67	1112	23.7	802	6 ABO25212	Abo25212 Novel hum
68	1112	23.7	802	6 ABO72218	Abo72218 Novel hum
69	1112	23.7	802	6 ABO48489	Abo48489 Human sec
70	1112	23.7	802	6 ABO61096	Abo61096 Human PRO
71	1112	23.7	802	6 ABO80365	Abo80365 Human sec
72	1112	23.7	802	6 ADA24708	Ada24708 Novel hum
73	1112	23.7	802	6 ABO19667	Abo19667 Novel hum
74	1112	23.7	802	6 ADA12369	Ada12369 Human sec
75	1112	23.7	802	6 ABO19558	Abo19558 Novel hum
76	1112	23.7	802	7 ADB73675	Adb73675 Human PRO
77	1112	23.7	802	7 ADB76391	Adb76391 Human PRO
78	1112	23.7	802	7 ADC43817	Adc43817 Human sec
79	1112	23.7	802	7 ADC61577	Adc61577 Human sec
80	1112	23.7	802	7 ADC63541	Adc63541 Human sec
81	1112	23.7	802	7 ADC66641	Adc66641 Human sec
82	1112	23.7	802	7 ADC68765	Adc68765 Human sec
83	1112	23.7	802	7 ADC68825	Adc68825 Human sec
84	1112	23.7	802	7 ADC67890	Adc67890 Human sec
85	1112	23.7	802	7 ADC41210	Adc41210 Human sec
86	1112	23.7	802	7 ADC67265	Adc67265 Human sec
87	1112	23.7	802	7 ADC62201	Adc62201 Human sec
88	1112	23.7	802	7 ADC41834	Adc41834 Human sec
89	1112	23.7	802	7 ADB49203	Adb49203 Human sec
90	1112	23.7	802	7 ADB35257	Adb35257 Human sec
91	1112	23.7	802	7 ADE16371	Ade16371 Human sec
92	1112	23.7	802	7 ADD72986	Add72986 Human sec
93	1112	23.7	802	7 ADD72344	Add72344 Human sec
94	1112	23.7	802	7 ADE16995	Ade16995 Human sec
95	1112	23.7	802	7 ADF47009	Adf47009 Human sec
96	1112	23.7	802	7 ADG52766	Adg52766 Human sec
97	1112	23.7	802	7 ADG60086	Adg60086 Human sec
98	1112	23.7	802	7 ADI60846	Adi60846 Human sec

ALIGNMENTS				Ade48503 Human sec	
				Ade89604 Human sec	
99	1112	23.7	802 8	ADE48503	
100	1112	23.7	802 8	ADE89604	
RESULT 1					
AAB19552					
ID	AAB19552 standard; protein; 855 AA.				
XX					
AC	AAB19552;				
XX					
DT	22-JAN-2001 (first entry)				
XX					
DE	Human matriptase.				
XX					
KM	Matriptase; serine protease; human; breast cancer; pre-malignancy;				
KM	actinic keratosis; leukoplakia; Barrett's epithelium;				
KM	columnar metaplasia; ulcerative colitis; bowenoid papulosis;				
KM	adenomatous colorectal polyp; Oesophag erythroplasia;				
KM	vulvar intraepithelial neoplasia; tumour; metastasis; therapy.				
XX					
OS	Homo sapiens.				
XX					
FH	Location/Qualifiers				
FT	Key				
FT	Region				
FT	Peptide				
FT	Modified-site				
FT	Region				
FT	Region				
FT	Region				
FT	Region				
FT	Domain				
FT	Cleavage-site				
XX					
PN	W0200053232-A1.				
XX					
PD	14-SEP-2000.				
XX					
PF	10-MAR-2000; 2000MO-US006111.				
XX					
PR	12-MAR-1999; 99US-0124006P.				
XX					
PA	(GEOU) UNIV GEORGETOWN.				
XX					
PI	Dickson RB, Lin C, Johnson M, Wang S, Enyedy I;				
XX					
DR	WPI; 2000-594268/56.				
DR	N-PSDB; AAA88493.				
XX					
PT	Treating malignancies, premalignant and pathologic conditions in a				
PT	subject, comprises administering matriptase modulating agent.				
XX					
PS	Claim 14; Fig 16; 11pp; English.				
XX					
CC	The present sequence is that of the full-length form of human matriptase,				
CC	a trypsin-like protease. This has an additional 172 amino acids compared				
CC	with the truncated form of matriptase given in AAB19551. Either form can				
CC	be produced in transformed or transfected cells using a claimed method.				

CC	The zymogen (inactive) form of matriptase is a single-chain protein. The				
CC	active 2-chain form strongly interacts with fragments of a Kunitz-type				
CC	serine protease inhibitor (hepatocyte growth factor activator inhibitor,				
CC	HAI-1) to form SDS-stable complexes. In breast cancer cells, matriptase				
CC	is present mainly as the uncomplexed form. Only the complexed matriptase				
CC	in detected in human milk. The invention is directed to a method of				
CC	detecting a malignancy or a pre-malignant lesion in breast or other				
CC	tissue by detecting the presence of single- or 2-chain forms of				
CC	matriptase in the tissue. The object is to inhibit tumour onset, tumour				
CC	growth and metastasis. Malignancies and pre-malignant conditions				
CC	characterised by expression of the zymogen or activated form of				
CC	matriptase are treated by administering an inhibitor of matriptase,				
CC	especially a Bowman-Birk inhibitor. The pre-malignant condition is				
CC	atypical ductal hyperplasia of the breast, actinic keratosis,				
CC	leukoplakia, Barrett's epithelium of the esophagus, ulcerative colitis,				
CC	adenomatous colorectal polyps, erythroplasia of the Queyrat, Bowen's				
CC	disease, bowenoid papulosis, vulvar intraepithelial neoplasia or				
CC	dysplastic changes to the cervix. The invention also provides methods for				
CC	in vivo or in vitro diagnosis of malignancy or pre-malignant lesion, and				
CC	methods of identifying matriptase modulators, including activators and				
CC	inhibitors				
XX					
SQ	Sequence 855 AA:				
Query Match	99.5%;	Score 4659;	DB 3;	Length 855;	
Best Local Similarity	99.6%;	Pred. No. 9.4e-308;			
Matches 852;	Conservative	0;	Mismatches 3;	Indels	0; Gaps 0;
QY	1 MSGRARRKGGGPKDGGAGLKYNRRHKNVGLSEGVPLPVNNVKKYKRGPRWVLLAA 60				
DB	1 MSGRARRKGGGPKDGGAGLKYNRRHKNVGLSEGVPLPVNNVKKYKRGPRWVLLAA 60				
QY	61 VLIGLLVLLIGLGFVLMHLQYRDVRYQKYNKGYMTITNENFVDAYENNSTEVSLSKY 120				
DB	61 VLIGLLVLLIGLGFVLMHLQYRDVRYQKYNKGYMTITNENFVDAYENNSTEVSLSKY 120				
QY	121 KDLAKLLYSGVPLPGYHKESAVTAPSEGSVINYNSERSIPDLYBEARVMAERVVM 180				
DB	121 KDLAKLLYSGVPLPGYHKESAVTAPSEGSVINYNSERSIPDLYBEARVMAERVVM 180				
QY	181 LPPRRSLKSFVTVSVVAPPTDSKTQVORTQDNCSCGLHARGVLMRFPTTGGPDSPYPA 240				
DB	181 LPPRRSLKSFVTVSVVAPPTDSKTQVORTQDNCSCGLHARGVLMRFPTTGGPDSPYPA 240				
QY	241 HARQWALRGDADSVSLTFRSPDLASCDEGSDLVTVNTLSPMPEHALVOLCGTYPSP 300				
DB	241 HARQWALRGDADSVSLTFRSPDLASCDEGSDLVTVNTLSPMPEHALVOLCGTYPSP 300				
QY	301 YNLTFRSSQNVLLITLITNTERHHPGFATFPLPMWSSCGGLRRAQGTFFNPGYGH 360				
DB	301 YNLTFRSSQNVLLITLITNTERHHPGFATFPLPMWSSCGGLRRAQGTFFNPGYGH 360				
QY	361 PNVIDCTWNIENVNNOHVKKRFKFLYLBEGVAGTCPODYVINGEKYCGESQFVVT 420				
DB	361 PNVIDCTWNIENVNNOHVKKRFKFLYLBEGVAGTCPODYVINGEKYCGESQFVVT 420				
QY	421 NSNKITVRFSHDSQYTDGTFLAETLSYDSDPCPGQFTCRGRCIRKEMLCDMADCTDH 480				
DB	421 NSNKITVRFSHDSQYTDGTFLAETLSYDSDPCPGQFTCRGRCIRKEMLCDMADCTDH 480				
QY	481 SDELNCSGDAHQFTCKNFKCPLFWVCDSVNDCCGNSDQGCSCPAQTFRCGNGKLSK 540				
DB	481 SDELNCSGDAHQFTCKNFKCPLFWVCDSVNDCCGNSDQGCSCPAQTFRCGNGKLSK 540				
QY	541 SQOCNGKDDCGSDSDASCPKVNVVCTKHTYRCLNGLCLSKNPECDGSDSDSDSK 600				
DB	541 SQOCNGKDDCGSDSDASCPKVNVVCTKHTYRCLNGLCLSKNPECDGSDSDSDSK 600				
QY	601 DCCGGLRSFTQARVVGTDADGEMPMQVSLHALQGHICGASLISPMNLVSAACYYID 660				
DB	601 DCCGGLRSFTQARVVGTDADGEMPMQVSLHALQGHICGASLISPMNLVSAACYYID 660				
QY	661 DRGFRYSDPFTQWTAFLGLHDQSORSAPOVERLKRILISHPFPNDFTFYDIALBLEXP 720				

```

Db 661 DRGRYSDPQTQWTAFLGLHDQSRSAFCVQERLKRISHPFNFDFDYDIALLELEKP 720
Qy 721 AEYSMTWPTICLPDASHVFPAGKAIWWTGNGHTOYGGTGAIILOKGEIRVYNQTTCEML 780
Db 721 AEYSMTWPTICLPDASHVFPAGKAIWWTGNGHTOYGGTGAIILOKGEIRVYNQTTCEML 780
Qy 781 PQQITPRMNCVGFSLSGVDSGQDSCGGLSSVEADGRIFGAGVYVWGDGAGRNKPGVYT 840
Db 781 PQQITPRMNCVGFSLSGVDSGQDSCGGLSSVEADGRIFGAGVYVWGDGAGRNKPGVYT 840
Qy 841 RLPLFRDWIKENTGV 855
Db 841 RLPLFRDWIKENTGV 855

RESULT 2
AAB35465
ID AAB35465 standard; protein; 855 AA.
AC AAB35465;
XX
XX 06-JUN-2001 (first entry)
XX
XX Human membrane-type serine protease MT-SPI.
XX
XX Human; membrane-type serine protease; MT-SPI; cancer.
XX
XX Homo sapiens.
XX
XX WO200123524-A2.
XX
XX PD 05-APR-2001.
XX
XX PF 02-OCT-2000; 2000MO-USO27250.
XX
XX PR 30-SEP-1999; 99US-00410362.
XX
XX PA (REGC ) UNIV CALIFORNIA.
XX
XX PI Craik CS, Takeuchi T, Shuman M;
XX
XX DR WPI; 2001-245002/25.
XX
XX PT N-PSDB; AAF28099.
XX
XX PT New nucleic acid encoding a membrane type serine protease, useful for the
XX diagnosis, prognosis and treatment of cancer, particularly metastatic
XX cancers.
XX
XX PS Claim 7; Fig 1; 102pp; English.
XX
XX CC The present invention provides the protein and coding sequences for the
XX novel human membrane-type serine protease MT-SPI. Increased expression of
XX this protein is associated with cancer, and so the sequences can be used
XX in cancer diagnosis and the identification of treatments. The present
XX sequence is the MT-SPI protein
XX
XX SQ Sequence 855 AA;
XX
XX Query Match 99.5%; Score 4659; DB 4; Length 855;
XX Best Local Similarity 99.6%; Pred. No. 9,4e-308;
XX Matches 852; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

Db 121 KDALKLYSGVPLGIPYHKESAVTAFSEGSVIAYWSEFSIPOHLVEAEARVMAEERVVM 180
Qy 181 LPPRARSLSKPVYTVSVVAFPTDSKTQVORTODNSCSFGLHARGVLAHFTTGPFPDPPYA 240
Db 181 LPPRARSLSKPVYTVSVVAFPTDSKTQVORTODNSCSFGLHARGVLAHFTTGPFPDPPYA 240
Qy 241 HARCQMLARDADSVLSLTFRSFDLASCDERGSIDLVTYNTLSMEPHALVOLCGTYP 300
Db 241 HARCQMLARDADSVLSLTFRSFDLASCDERGSIDLVTYNTLSMEPHALVOLCGTYP 300
Qy 301 YNLTFHSSQVNLTLITLNTERRHPGEATFPQIPRMSCGGRLRKAQGFNSPYTGYH 360
Db 301 YNLTFHSSQVNLTLITLNTERRHPGEATFPQIPRMSCGGRLRKAQGFNSPYTGYH 360
Qy 361 PPNIDCTWNLVPPNQHVKVRFKFEYLLBGPVAGTCPKDYVEINRKYCGERSQFVYTS 420
Db 361 PPNIDCTWNLVPPNQHVKVRFKFEYLLBGPVAGTCPKDYVEINRKYCGERSQFVYTS 420
Qy 421 NSNKITVRPHSDSYTDYGFIAEYLSYDSSDPCPGQFTCRTRGCIKRLRCDGADCTDH 480
Db 421 NSNKITVRPHSDSYTDYGFIAEYLSYDSSDPCPGQFTCRTRGCIKRLRCDGADCTDH 480
Qy 481 SDEINCSDAQHOFCKNKFCKPLFWVCDSDVNDGDNSSDEQSCSPAQTRFCSNGKCLSK 540
Db 481 SDEINCSDAQHOFCKNKFCKPLFWVCDSDVNDGDNSSDEQSCSPAQTRFCSNGKCLSK 540
Qy 541 SQQNGKDDCGDSDSCRPVNVVCTKHTYRCLNGLCLSKNPECDGKEDSDSDSDEK 600
Db 541 SQQNGKDDCGDSDSCRPVNVVCTKHTYRCLNGLCLSKNPECDGKEDSDSDSDEK 600
Qy 601 DCDGGLRSFTROARVVGTDADDEGEMPMQVSLHALGQGHICGASLISPNMLVSAHICYID 660
Db 601 DCDGGLRSFTROARVVGTDADDEGEMPMQVSLHALGQGHICGASLISPNMLVSAHICYID 660
Qy 661 DRGRYSDPQTQWTAFLGLHDQSRSAFCVQERLKRISHPFNFDFDYDIALLELEKP 720
Db 661 DRGRYSDPQTQWTAFLGLHDQSRSAFCVQERLKRISHPFNFDFDYDIALLELEKP 720
Qy 721 AEYSMTWPTICLPDASHVFPAGKAIWWTGNGHTOYGGTGAIILOKGEIRVYNQTTCEML 780
Db 721 AEYSMTWPTICLPDASHVFPAGKAIWWTGNGHTOYGGTGAIILOKGEIRVYNQTTCEML 780
Qy 781 PQQITPRMNCVGFSLSGVDSGQDSCGGLSSVEADGRIFGAGVYVWGDGAGRNKPGVYT 840
Db 781 PQQITPRMNCVGFSLSGVDSGQDSCGGLSSVEADGRIFGAGVYVWGDGAGRNKPGVYT 840
Qy 841 RLPLFRDWIKENTGV 855
Db 841 RLPLFRDWIKENTGV 855

RESULT 3
AD116817
ID AD116817 standard; protein; 855 AA.
XX
XX AC AD116817;
XX
XX DT 15-APR-2004 (first entry)
XX
XX DE Human NOXV protein homologue SegID 353.
XX
XX KW human; NOXV; cardiomyopathy; atherosclerosis; cancer; diabetes;
XX inflammation; autoimmune disorder; allergy; blood disorder;
XX acquired immunodeficiency syndrome; AIDS; obesity; ascitima;
XX immunoglobulin (Ig)A nephropathy; cirrhosis; arthritis;
XX Alzheimer's disease; infection; str.
XX
XX OS Homo sapiens.
XX
XX PN WO200268649-A2.
XX
XX PD 06-SEP-2002.
XX

```

PF 31-JAN-2002; 2002W0-US002785.
XX
PR 31-JAN-2001; 2001US-0265395P.
PR 31-JAN-2001; 2001US-0265412P.
PR 31-JAN-2001; 2001US-0265514P.
PR 31-JAN-2001; 2001US-0265517P.
PR 02-FEB-2001; 2001US-0265406P.
PR 05-FEB-2001; 2001US-0266767P.
PR 07-FEB-2001; 2001US-0266975P.
PR 07-FEB-2001; 2001US-0267057P.
PR 08-FEB-2001; 2001US-0267459P.
PR 09-FEB-2001; 2001US-0267823P.
PR 15-FEB-2001; 2001US-0268974P.
PR 26-FEB-2001; 2001US-0271664P.
PR 27-FEB-2001; 2001US-0271839P.
PR 27-FEB-2001; 2001US-0271855P.
PR 02-MAR-2001; 2001US-0272788P.
PR 02-MAR-2001; 2001US-0273046P.
PR 14-MAR-2001; 2001US-0275925P.
PR 14-MAR-2001; 2001US-0275947P.
PR 14-MAR-2001; 2001US-0275950P.
PR 14-MAR-2001; 2001US-0275989P.
PR 15-MAR-2001; 2001US-0276448P.
PR 15-MAR-2001; 2001US-0276450P.
PR 16-MAR-2001; 2001US-0276397P.
PR 16-MAR-2001; 2001US-0276768P.
PR 20-MAR-2001; 2001US-0278652P.
PR 26-MAR-2001; 2001US-0278775P.
PR 26-MAR-2001; 2001US-0278778P.
PR 29-MAR-2001; 2001US-0279882P.
PR 29-MAR-2001; 2001US-0279884P.
PR 30-MAR-2001; 2001US-0280147P.
PR 11-APR-2001; 2001US-0282992P.
PR 11-APR-2001; 2001US-0283083P.
PR 20-APR-2001; 2001US-0285133P.
PR 23-APR-2001; 2001US-0285749P.
PR 03-MAY-2001; 2001US-0286327P.
PR 03-MAY-2001; 2001US-0288504P.
PR 29-MAY-2001; 2001US-0294047P.
PR 30-MAY-2001; 2001US-0294473P.
PR 08-JUN-2001; 2001US-0296644P.
PR 18-JUN-2001; 2001US-0298959P.
PR 19-JUN-2001; 2001US-0299324P.
PR 13-AUG-2001; 2001US-0312020P.
PR 16-AUG-2001; 2001US-0312889P.
PR 16-AUG-2001; 2001US-0312908P.
PR 21-AUG-2001; 2001US-0313390P.
PR 28-AUG-2001; 2001US-0315470P.
PR 31-AUG-2001; 2001US-0316447P.
PR 07-SEP-2001; 2001US-0318115P.
PR 07-SEP-2001; 2001US-0318118P.
PR 12-SEP-2001; 2001US-0318740P.
PR 19-SEP-2001; 2001US-0323379P.
PR 18-OCT-2001; 2001US-0330245P.
PR 18-OCT-2001; 2001US-0330308P.
PR 14-NOV-2001; 2001US-0332701P.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Therneer VT, Spytek KA, Zernhuen BD, Patturajan M, Shinkets RA,
PI Li L, Gangolli EA, Padigar M, Anderson DM, Raetelli L, Miller CE;
PI Gurelch VL, Tappier RJ, Gusev VY, Colman SD, Wolenc AR, Pena CZA;
PI Furtak K, Grosse WM, Alsobrook JP, Lepley SM, Rieger DK, Burgess CE;
XX
DR WPI; 2002-706998/76.
XX
PT New NOVX polypeptides and nucleic acids, useful for preventing or
PT treating NOVX-associated disorders, e.g. cancer, cardiomyopathy,
PT atherosclerosis, or diabetes, and in chromosome mapping, tissue typing or
PT pharmacogenomics.
XX
PS Disclosure; SEQ ID NO 353; 1498pp; English.

CC This invention relates to a novel nucleic acids, and encoded polypeptides
CC thereof, which have properties related to the stimulation of biochemical
CC or physiological responses in a cell, tissue, organ or organism.
CC Specifically, it refers to the use of biologically active fragments for
CC diagnostic and prognostic assays and furthermore in the treatment of
CC diverse pathological conditions. The present invention describes novel
CC human and murine NOVX proteins, as well as methods to modulate their
CC expression using antisense oligos, ribozymes and peptide nucleic acids.
CC The NOVX polypeptides, polynucleotides and antibodies are useful in
CC treating or preventing NOVX-associated disorders, e.g. cardiomyopathy,
CC atherosclerosis, cancer and diabetes. Furthermore, they may be used in
CC treating or preventing diseases such as inflammation, autoimmune
CC disorders, allergies, blood disorders, acquired immunodeficiency syndrome
CC (AIDS), obesity, asthma, immunoglobulin (Ig) A nephropathy, cirrhosis,
CC arthritis, Alzheimer's disease, infections, stroke, muscular dystrophy
CC and epilepsy. Accordingly, these molecules have many activities including
CC cytostatic, cardiant, antiinflammatory, immunosuppressive, antiallergic,
CC hemostatic, anti-HIV, antidiabetic, antihypertensive, anorectic,
CC neuroprotective, neurotropic, antirheumatic, hepatotropic,
CC neuroprotective, neurotropic, antibacterial, virucide, antiparasitic,
CC relaxant and anticonvulsant. In addition, they are useful in screening
CC assays to identify small molecules that modulate or inhibit, for example,
CC neurogenesis, wound healing and angiogenesis. The nucleic acids are also
CC used as in chromosome mapping, tissue typing, preventive medicine and
CC pharmacogenomics. This polypeptide is a homologue of a human NOVX protein
CC of the invention.
XX
SQ Sequence 855 AA;
Query Match 99.5%; Score 4659; DB 5; Length 855;
Best Local Similarity 99.0%; Pred. No. 9.4e-308;
Matches 852; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 MGSDRRKGGGPKDAGLKYNSRHEKVGLEGEVFLPVNNVKKVKGKGRWVLLAA 60
DB 1 MGSDRRKGGGPKDAGLKYNSRHEKVGLEGEVFLPVNNVKKVKGKGRWVLLAA 60
QY VILGLILVILGIFLVYHLYQYRDVYQKXKNGYRITNENFVDAYENSSTEVSLASKY 120
DB VILGLILVILGIFLVYHLYQYRDVYQKXKNGYRITNENFVDAYENSSTEVSLASKY 120
QY 121 KDLAKLLYSQVPLGPGYHKEASVTAESGSLVIAVYSEPSIPQHLVEAEKRVAAERVVM 180
DB 121 KDLAKLLYSQVPLGPGYHKEASVTAESGSLVIAVYSEPSIPQHLVEAEKRVAAERVVM 180
QY 121 KDLAKLLYSQVPLGPGYHKEASVTAESGSLVIAVYSEPSIPQHLVEAEKRVAAERVVM 180
DB 121 KDLAKLLYSQVPLGPGYHKEASVTAESGSLVIAVYSEPSIPQHLVEAEKRVAAERVVM 180
QY 181 LPPRASLSKSFVTSVVAPEPTDSKTQVQTQDNCSEGLHARGEVLMFTTGPDPSPYPA 240
DB 181 LPPRASLSKSFVTSVVAPEPTDSKTQVQTQDNCSEGLHARGEVLMFTTGPDPSPYPA 240
QY 241 HARCQALRGDADSVLSLFRSFDLASCDERGSDLVTVNTLSPMERHALVOLCGTYPSP 300
DB 241 HARCQALRGDADSVLSLFRSFDLASCDERGSDLVTVNTLSPMERHALVOLCGTYPSP 300
QY 241 HARCQALRGDADSVLSLFRSFDLASCDERGSDLVTVNTLSPMERHALVOLCGTYPSP 300
DB 241 HARCQALRGDADSVLSLFRSFDLASCDERGSDLVTVNTLSPMERHALVOLCGTYPSP 300
QY 301 YNLTFFSSQVNLITLITNTERHHPGEATFPQLPMSSCGGLRAQGFNSPYTGHY 360
DB 301 YNLTFFSSQVNLITLITNTERHHPGEATFPQLPMSSCGGLRAQGFNSPYTGHY 360
QY 361 PNVIDCTWNIENVNNOHVKRFKFLYLBEGVAGTCPOXYVINEKCYGERSQCVVTS 420
DB 361 PNVIDCTWNIENVNNOHVKRFKFLYLBEGVAGTCPOXYVINEKCYGERSQCVVTS 420
QY 421 NSNKITVRPHSDSYDTGFLAETLSYSSDPCEGQFTCTGRCIRKELRDGMADCTTH 480
DB 421 NSNKITVRPHSDSYDTGFLAETLSYSSDPCEGQFTCTGRCIRKELRDGMADCTTH 480
QY 481 SDELNCSCAQHOFCTKKNKFCRPLFWVCSVNDGNSBQGSCEPAQTFRSGNGLSK 540
DB 481 SDELNCSCAQHOFCTKKNKFCRPLFWVCSVNDGNSBQGSCEPAQTFRSGNGLSK 540
QY 541 SQQCNKDKDCGSDSDASCPKVVVYCTGHTYRCNLGLCLSKGNPCDGEDSDSDSK 600
DB 541 SQQCNKDKDCGSDSDASCPKVVVYCTGHTYRCNLGLCLSKGNPCDGEDSDSDSK 600

QY 601 DCDGSLRFTFOARVVGSTADDEGEMPMQVSLHALGCHICGASLISPMVLVSAHACYID 660
Db 601 DCDGSLRFTFOARVVGSTADDEGEMPMQVSLHALGCHICGASLISPMVLVSAHACYID 660
QY 661 DRGRYSADPTQMTAFGLIHDSQSRAPQVQRRLKR11SHDPFDFDYDIALLELEKP 720
Db 661 DRGRYSADPTQMTAFGLIHDSQSRAPQVQRRLKR11SHDPFDFDYDIALLELEKP 720
QY 721 AEVSMRPFICLPASHVFPFGAKAIWYGMGHTQYGGALILQKGEIRVINTOTCEML 780
Db 721 AEVSMRPFICLPASHVFPFGAKAIWYGMGHTQYGGALILQKGEIRVINTOTCEML 780
QY 781 PQDITPRMVCVGFSGVDSQGDSPGSPSSVEADGRIFGAGVYMWGDCAGRNKPGVYT 840
Db 781 PQDITPRMVCVGFSGVDSQGDSPGSPSSVEADGRIFGAGVYMWGDCAGRNKPGVYT 840
QY 841 RLPLFRDMIKENTGV 855
Db 841 RLPLFRDMIKENTGV 855
RESULT 4
AD116883
ID AD116883 standard; protein; 855 AA.
XX
AC AD116883;
XX
DT 15-APR-2004 (first entry)
XX
DE Human NOVX protein homologue seqid 419.
XX
KW human; NOVX; cardiomyopathy; atherosclerosis; cancer; diabetes;
KW inflammation; autoimmune disorder; allergy; blood disorder;
KW acquired immunodeficiency syndrome; AIDS; obesity; asthma;
KW immunoglobulin (Ig)A nephropathy; cirrhosis; arthritis;
KW Alzheimer's disease; infection; str.
XX
OS Homo sapiens.
XX
PN W0200268649-A2.
XX
PD 06-SEP-2002.
XX
PF 31-JAN-2002; 2002W0-US002785.
XX
PR 31-JAN-2001; 2001US-0265395P.
PR 31-JAN-2001; 2001US-0265412P.
PR 31-JAN-2001; 2001US-0265514P.
PR 31-JAN-2001; 2001US-0265517P.
PR 02-FEB-2001; 2001US-0266406P.
PR 05-FEB-2001; 2001US-0266767P.
PR 07-FEB-2001; 2001US-0266975P.
PR 07-FEB-2001; 2001US-0267057P.
PR 08-FEB-2001; 2001US-0267459P.
PR 09-FEB-2001; 2001US-0267823P.
PR 15-FEB-2001; 2001US-0268974P.
PR 26-FEB-2001; 2001US-0271664P.
PR 27-FEB-2001; 2001US-0271839P.
PR 02-MAR-2001; 2001US-0272788P.
PR 02-MAR-2001; 2001US-0273046P.
PR 14-MAR-2001; 2001US-0275925P.
PR 14-MAR-2001; 2001US-0275947P.
PR 14-MAR-2001; 2001US-0275950P.
PR 15-MAR-2001; 2001US-0275989P.
PR 15-MAR-2001; 2001US-0276448P.
PR 15-MAR-2001; 2001US-0276450P.
PR 16-MAR-2001; 2001US-0276397P.
PR 16-MAR-2001; 2001US-0276768P.
PR 20-MAR-2001; 2001US-0278652P.
PR 26-MAR-2001; 2001US-0278775P.
PR 26-MAR-2001; 2001US-0278778P.
PR 29-MAR-2001; 2001US-0279882P.

PR 29-MAR-2001; 2001US-0279884P.
PR 30-MAR-2001; 2001US-0280147P.
PR 11-APR-2001; 2001US-0282992P.
PR 11-APR-2001; 2001US-0283063P.
PR 20-APR-2001; 2001US-0285133P.
PR 23-APR-2001; 2001US-0285749P.
PR 03-MAY-2001; 2001US-0288372P.
PR 03-MAY-2001; 2001US-0288504P.
PR 29-MAY-2001; 2001US-0294047P.
PR 30-MAY-2001; 2001US-0294473P.
PR 08-JUN-2001; 2001US-0296964P.
PR 18-JUN-2001; 2001US-0298959P.
PR 19-JUN-2001; 2001US-0299324P.
PR 13-AUG-2001; 2001US-0312020P.
PR 16-AUG-2001; 2001US-0312889P.
PR 16-AUG-2001; 2001US-0312908P.
PR 21-AUG-2001; 2001US-0313390P.
PR 28-AUG-2001; 2001US-0315470P.
PR 31-AUG-2001; 2001US-0316447P.
PR 07-SEP-2001; 2001US-0318115P.
PR 12-SEP-2001; 2001US-0318118P.
PR 12-SEP-2001; 2001US-0318740P.
PR 19-SEP-2001; 2001US-0323379P.
PR 18-OCT-2001; 2001US-0320245P.
PR 18-OCT-2001; 2001US-0330308P.
PR 14-NOV-2001; 2001US-0332701P.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Tchernev VT, Spytek RA, Zernhusen BD, Paturajan M, Shinkets RA,
XX Li L, Gangoli EA, Padigaru M, Anderson DW, Rascelli L, Miller CE,
XX Gerlach VL, Taupier RJ, Gusev VY, Colman SD, Wolsen AR, Pena CRA,
XX Furtak K, Grosse WM, Alsbrook JP, Lepley DM, Rieger DK, Burgess CE,
XX WPI; 2002-706998/76.
XX
XX New NOVX polypeptides and nucleic acids, useful for preventing or
XX treating NOVX-associated disorders, e.g. cancer, cardiomyopathy,
XX atherosclerosis, or diabetes, and in chromosome mapping, tissue typing or
XX pharmacogenomics.
XX
XX Disclosure; SEQ ID NO 419; 1498pp; English.
XX
XX This invention relates to a novel nucleic acids, and encoded polypeptides
XX thereof, which have properties related to the stimulation of biochemical
XX or physiological responses in a cell, tissue, organ or organism.
XX Specifically, it refers to the use of biologically active fragments for
XX diagnostic and prognostic assays and furthermore in the treatment of
XX diverse pathological conditions. The present invention describes novel
XX human and murine NOVX proteins, as well as methods to modulate their
XX expression using antisense oligos, ribozymes and peptide nucleic acids.
XX The NOVX polypeptides, polynucleotides and antibodies are useful in
XX treating or preventing NOVX-associated disorders, e.g. cardiomyopathy,
XX atherosclerosis, cancer and diabetes. Furthermore, they may be used in
XX treating or preventing diseases such as inflammation, autoimmune
XX disorders, allergies, blood disorders, acquired immunodeficiency syndrome
XX (AIDS), obesity, asthma, immunoglobulin (Ig)A nephropathy, cirrhosis,
XX arthritis, Alzheimer's disease, infections, stroke, muscular dystrophy
XX and epilepsy. Accordingly, these molecules have many activities including
XX cytoskeletal, cardiac, antiinflammatory, immunosuppressive, antiallergic,
XX haemostatic, anti-HIV, antidiabetic, antiatherosclerotic, anorectic,
XX antiasthmatic, nephrotropic, antiarthritic, hepatotropic,
XX neuroprotective, neurotropic, antibacterial, vitruide, antiparasitic,
XX relaxant and anticonvulsant. In addition, they are useful in screening
XX assays to identify small molecules that modulate or inhibit, for example,
XX neurogenesis, wound healing and angiogenesis. The nucleic acids are also
XX used as in chromosome mapping, tissue typing, preventive medicine and
XX pharmacogenomics. This polypeptide is a homologue of a human NOVX protein
XX of the invention.
XX
XX Sequence 855 AA;
XX
XX Query Match 99.5%; Score 4659; DB 5; Length 855;

Best Local Similarity 99.6%; Pred. No. 9.4e-308;
Matches 852; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
QY 1 MGSDRRKGGGGPPDPGAGLKYNSRHKVNGLEGEVFLPNNVKKYKXKPGGWVLA 60
Db 1 MGSDRRKGGGGPPDPGAGLKYNSRHKVNGLEGEVFLPNNVKKYKXKPGGWVLA 60
QY VLIQLLVLIIGLIGFLVHMLQYRDVVOVKVKNYKRIINENFVDAEENSNTSEFVSLASKV 120
Db VLIQLLVLIIGLIGFLVHMLQYRDVVOVKVKNYKRIINENFVDAEENSNTSEFVSLASKV 120
QY 121 KDAKLKLYSGVPFLGPRHKESAATAFSEGSYIATYNSBFSIPQHLVEAEKRVMAEERYVM 180
Db 121 KDAKLKLYSGVPFLGPRHKESAATAFSEGSYIATYNSBFSIPQHLVEAEKRVMAEERYVM 180
QY 121 KDAKLKLYSGVPFLGPRHKESAATAFSEGSYIATYNSBFSIPQHLVEAEKRVMAEERYVM 180
Db 121 KDAKLKLYSGVPFLGPRHKESAATAFSEGSYIATYNSBFSIPQHLVEAEKRVMAEERYVM 180
QY 181 LPPRARSLSKSVVYSVVAFPDTSKTVORTODNSCSFGILHARGVLAEMPTTGPFPDSYPA 240
Db 181 LPPRARSLSKSVVYSVVAFPDTSKTVORTODNSCSFGILHARGVLAEMPTTGPFPDSYPA 240
QY 241 HARCQMALRGDADSVLSLTFRSFDLASCDEKSGDLVTYNTLSPMERHALVOLCGTYPPS 300
Db 241 HARCQMALRGDADSVLSLTFRSFDLASCDEKSGDLVTYNTLSPMERHALVOLCGTYPPS 300
QY 301 YNLTFFSSQNVLLITLTNTERRHGPBATTFOQLPRMSSCGGRLRKAQGTFFNSYYRGHY 360
Db 301 YNLTFFSSQNVLLITLTNTERRHGPBATTFOQLPRMSSCGGRLRKAQGTFFNSYYRGHY 360
QY 361 PPNIDCTMNIEVPNNQHVKKRFKFFYLLEPGVATGCPKQVVEINGEKYCGEERQFVYTS 420
Db 361 PPNIDCTMNIEVPNNQHVKKRFKFFYLLEPGVATGCPKQVVEINGEKYCGEERQFVYTS 420
QY 421 NSNKITVRFHSDQSYDTGTGLAEVLSYSDSPCPGQFTCTGRCIRKELRCGWDCTDH 480
Db 421 NSNKITVRFHSDQSYDTGTGLAEVLSYSDSPCPGQFTCTGRCIRKELRCGWDCTDH 480
QY 481 SDEINCSGDAGHOTCTCKNKKFKPLFWICDSVNDGSDNSDEGSCCPAQTFRCSNGKCLSK 540
Db 481 SDEINCSGDAGHOTCTCKNKKFKPLFWICDSVNDGSDNSDEGSCCPAQTFRCSNGKCLSK 540
QY 541 SQQNGKDDCGDSDDEASCPKVNVTCTKTYRCLANGCLSKNGPECDGKEDCSGSDSK 600
Db 541 SQQNGKDDCGDSDDEASCPKVNVTCTKTYRCLANGCLSKNGPECDGKEDCSGSDSK 600
QY 601 DCDGGLSFTROARVVGCTDADEGEPMQVSLHALGQGHICGASLLSPNMLVSAHGYID 660
Db 601 DCDGGLSFTROARVVGCTDADEGEPMQVSLHALGQGHICGASLLSPNMLVSAHGYID 660
QY 661 DRGRYSDPTQWTAFLGLHDQSASAPGVORRLKRIISHPFNDFTFDVIALLELEKP 720
Db 661 DRGRYSDPTQWTAFLGLHDQSASAPGVORRLKRIISHPFNDFTFDVIALLELEKP 720
QY 721 AEYSMWRTICLPDASHVFPAGKAIWYTGWCHTOYGGTGALILQKGEIRVINOTTENLL 780
Db 721 AEYSMWRTICLPDASHVFPAGKAIWYTGWCHTOYGGTGALILQKGEIRVINOTTENLL 780
QY 781 PQQITPRMWCVGLSGVDSCQDGSGLPSVEADGRIFGAGVVSWDGCAGRNKPQVYT 840
Db 781 PQQITPRMWCVGLSGVDSCQDGSGLPSVEADGRIFGAGVVSWDGCAGRNKPQVYT 840
QY 841 RLPLFRMWIKENTGV 855
Db 841 RLPLFRMWIKENTGV 855
```

RESULT 5

AD116876 standard; protein; 855 AA.

AD116876;

15-APR-2004 (first entry)

Human NOXV protein homologue SegID 412.

XX human; NOXV; cardiomyopathy; atherosclerosis; cancer; diabetes;
KW inflammation; autoimmune disorder; allergy; blood disorder;
KW acquired immunodeficiency syndrome; AIDS; obesity; asthma;
KW immunoglobulin (Ig)A nephropathy; cirrhosis; arthritis;
KW Alzheimer's disease; infection; str.

OS Homo sapiens.

PN W0200268649-A2.

PD 06-SEP-2002.

PF 31-JAN-2002; 2002W0-US002785.

PR 31-JAN-2001; 2001US-0265395P.

PR 31-JAN-2001; 2001US-0265412P.

PR 31-JAN-2001; 2001US-0265514P.

PR 02-FEB-2001; 2001US-0265517P.

PR 05-FEB-2001; 2001US-0266406P.

PR 07-FEB-2001; 2001US-0266767P.

PR 07-FEB-2001; 2001US-0266975P.

PR 08-FEB-2001; 2001US-0267057P.

PR 09-FEB-2001; 2001US-0267459P.

PR 15-FEB-2001; 2001US-0267823P.

PR 26-FEB-2001; 2001US-0271664P.

PR 27-FEB-2001; 2001US-0271839P.

PR 02-MAR-2001; 2001US-0271855P.

PR 14-MAR-2001; 2001US-0272788P.

PR 14-MAR-2001; 2001US-0273046P.

PR 14-MAR-2001; 2001US-0275925P.

PR 14-MAR-2001; 2001US-0275947P.

PR 14-MAR-2001; 2001US-0275950P.

PR 14-MAR-2001; 2001US-0275989P.

PR 15-MAR-2001; 2001US-0276448P.

PR 15-MAR-2001; 2001US-0276450P.

PR 16-MAR-2001; 2001US-0276397P.

PR 16-MAR-2001; 2001US-0276768P.

PR 20-MAR-2001; 2001US-0278652P.

PR 26-MAR-2001; 2001US-0278775P.

PR 26-MAR-2001; 2001US-0278778P.

PR 29-MAR-2001; 2001US-0279682P.

PR 29-MAR-2001; 2001US-0279684P.

PR 30-MAR-2001; 2001US-0280147P.

PR 11-APR-2001; 2001US-0282992P.

PR 11-APR-2001; 2001US-0283083P.

PR 20-APR-2001; 2001US-0285133P.

PR 23-APR-2001; 2001US-0285749P.

PR 03-MAY-2001; 2001US-0288327P.

PR 03-MAY-2001; 2001US-0288504P.

PR 29-MAY-2001; 2001US-0294047P.

PR 30-MAY-2001; 2001US-0294473P.

PR 08-JUN-2001; 2001US-0296364P.

PR 18-JUN-2001; 2001US-0298959P.

PR 19-JUN-2001; 2001US-0299324P.

PR 13-AUG-2001; 2001US-0312020P.

PR 16-AUG-2001; 2001US-0312889P.

PR 16-AUG-2001; 2001US-0312908P.

PR 21-AUG-2001; 2001US-0313390P.

PR 28-AUG-2001; 2001US-0315470P.

PR 31-AUG-2001; 2001US-0316447P.

PR 07-SEP-2001; 2001US-0318115P.

PR 07-SEP-2001; 2001US-0318118P.

PR 12-SEP-2001; 2001US-0318740P.

PR 19-SEP-2001; 2001US-0323379P.

PR 18-OCT-2001; 2001US-0330245P.

PR 18-OCT-2001; 2001US-0330308P.

PR 14-NOV-2001; 2001US-0332701P.

(CURA-) CURAGEN CORP.

Tchernev VT, Spytek KA, Zernusen BD, Paturajan M, Shinkets RA;

PI Li L, Gangoli EA, Padigaru M, Anderson DM, Raselli L, Miller CE;
 PI Gerlach VL, Taupier RJ, Gusev VY, Colman SD, Wolenc AR, Pena CE;
 PI Furtak K, Grosse WM, Alsbrook JP, Lepley DM, Rieger DK, Burgess CE;
 XX WPI; 2002-706998/76.
 XX
 XX New NOVX polypeptides and nucleic acids, useful for preventing or
 PT treating NOVX-associated disorders, e.g. cancer, cardiomyopathy,
 PT atherosclerosis, or diabetes, and in chromosome mapping, tissue typing or
 PT pharmacogenomics.
 XX
 XX Disclosure; SEQ ID NO 412; 1498bp; English.
 XX
 CC This invention relates to a novel nucleic acid, and encoded polypeptides
 CC thereof, which have properties related to the stimulation of biochemical
 CC or physiological responses in a cell, tissue, organ or organism.
 CC Specifically, it refers to the use of biologically active fragments for
 CC diagnostic and prognostic assays and furthermore in the treatment of
 CC diverse pathological conditions. The present invention describes novel
 CC human and murine NOVX proteins, as well as methods to modulate their
 CC expression using antisense oligos, ribozymes and peptide nucleic acids.
 CC The NOVX polypeptides, polynucleotides and antibodies are useful in
 CC treating or preventing NOVX-associated disorders, e.g. cardiomyopathy,
 CC atherosclerosis, cancer and diabetes. Furthermore, they may be used in
 CC treating or preventing diseases such as inflammation, autoimmune
 CC disorders, allergies, blood disorders, acquired immunodeficiency syndrome
 CC (AIDS), obesity, asthma, immunoglobulin (Ig)A nephropathy, cirrhosis,
 CC arthritis, Alzheimer's disease, infections, stroke, muscular dystrophy,
 CC and epilepsy. Accordingly, these molecules have many activities including
 CC cytoskeletal, cardiac, anti-inflammatory, immunosuppressive, antiallergic,
 CC haemostatic, anti-HIV, antidiabetic, antierosive, anorectic,
 CC antiasthmatic, nephrotropic, antiarthritic, hepatotropic,
 CC neuroprotective, nootropic, antibacterial, virucide, antiparasitic,
 CC relaxant and anticonvulsant. In addition, they are useful in screening
 CC assays to identify small molecules that modulate or inhibit, for example,
 CC neurogenesis, wound healing and angiogenesis. The nucleic acids are also
 CC used as in chromosome mapping, tissue typing, preventive medicine and
 CC pharmacogenomics. This polypeptide is a homologue of a human NOVX protein
 CC of the invention.
 CC
 XX
 XX Sequence 855 AA;
 SQ
 Query Match 99.5%; Score 4659; DB 5; Length 855;
 Best Local Similarity 99.6%; Pred. No. 9,4e-308;
 Matches 855; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 421 NSNKTIVRPHSDSYTDTGFLAELYSYDSSDPGQGTCTGTCGRICRKLRCQWADCTDH 480
 DB 421 NSNKTIVRPHSDSYTDTGFLAELYSYDSSDPGQGTCTGTCGRICRKLRCQWADCTDH 480
 QY 481 SDELNCSGADAGHPTCKNKECKPLFWVCDSVNDGNSDQSCSPAQTRFCSNGKLSK 540
 DB 481 SDELNCSGADAGHPTCKNKECKPLFWVCDSVNDGNSDQSCSPAQTRFCSNGKLSK 540
 QY 541 SQQNGNDGDDGSDASCKRANVYTCTKATYRCLNGLCLSKNPECDGDCSDGDEK 600
 DB 541 SQQNGNDGDDGSDASCKRANVYTCTKATYRCLNGLCLSKNPECDGDCSDGDEK 600
 QY 601 DCDGSLSPTRQARVVGTDADGEMPMOYSLHALGQGHICGASLISPMNLVSAHGYID 660
 DB 601 DCDGSLSPTRQARVVGTDADGEMPMOYSLHALGQGHICGASLISPMNLVSAHGYID 660
 QY 661 DRGRYSPTQMTAFGLADQSORSAFGVQERRIKRIISHPFNDFTDYDIALLELEK 720
 DB 661 DRGRYSPTQMTAFGLADQSORSAFGVQERRIKRIISHPFNDFTDYDIALLELEK 720
 QY 721 AYSVMWRPCLDPASVFPAGKAIWTTGHTQYGGTGLIIQKGRIRYINOTTENLL 780
 DB 721 AYSVMWRPCLDPASVFPAGKAIWTTGHTQYGGTGLIIQKGRIRYINOTTENLL 780
 QY 781 POOTPRMVCVFLSGGVSDSCQDSSGGLSSVEADGRIFGAVVSMGDGAGRRKPGVYT 840
 DB 781 POOTPRMVCVFLSGGVSDSCQDSSGGLSSVEADGRIFGAVVSMGDGAGRRKPGVYT 840
 QY 841 RLPLFRDMIKENTGV 855
 DB 841 RLPLFRDMIKENTGV 855
 RESULT 6
 ID ADN39867 standard; protein; 855 AA.
 XX
 XX ADN39867;
 AC
 XX
 XX 17-JUN-2004 (first entry)
 DT
 XX
 XX Cancer/angiogenesis/fibrosis-related polypeptide, SEQ ID NO: C237.
 DE
 XX
 KW Human, differential expression; cancer; angiogenic disorder;
 KW fibrotic disorder; psoriasis; leukaemia; heart disease; atherosclerosis;
 KW inflammatory disease; autoimmune disease;
 KW retinal neovascularisation syndrome; scarring; uterine fibroid;
 KW detection; diagnosis; prognosis; drug screening; drug targeting;
 KW wound healing; contraception; cytoskeletal; cardiac; immunomodulatory;
 KW vulnertary; gene therapy; vaccine.
 KW
 XX Homo sapiens.
 OS
 XX
 XX MO2003042661-A2.
 FN
 XX
 PD 22-MAY-2003.
 XX
 XX 13-NOV-2002; 2002WC-US036810.
 PF
 XX
 XX 13-NOV-2001; 2001US-0350666P.
 PR 21-NOV-2001; 2001US-0332464P.
 PR 29-NOV-2001; 2001US-0334383P.
 PR 03-DEC-2001; 2001US-0353594P.
 PR 14-DEC-2001; 2001US-0340371P.
 PR 08-JAN-2002; 2002US-0347211P.
 PR 10-JAN-2002; 2002US-0347349P.
 PR 08-FEB-2002; 2002US-0355250P.
 PR 13-FEB-2002; 2002US-0356714P.
 PR 20-FEB-2002; 2002US-0359077P.
 PR 29-MAR-2002; 2002US-0368099P.
 PR 04-APR-2002; 2002US-0370110P.
 PR 12-APR-2002; 2002US-0372246P.

PR 05-JUN-2002; 2002US-0386614P.
PR 16-JUL-2002; 2002US-0396839P.
PR 22-JUL-2002; 2002US-0397775P.
PR 22-JUL-2002; 2002US-0397845P.
PR 09-SEP-2002; 2002US-0409450P.
XX
XX (EOSB-) EOS BIOTECHNOLOGY INC.
PI Afar D, Aziz N, Ginsburg WM, Gish KC, Glyme R, Hevezi PA;
PI Mack DH, Murray R, Watson SR, Wilson KE, Zlocnik A;
XX
XX WPI; 2003-468649/44.
DR N-PSDB; ADN39650.
XX
XX Determining the presence or absence of a pathological cell in a patient,
PT useful for diagnosing, prognosing or treating cancer, comprises detecting
PT a nucleic acid in a biological sample.
XX
XX Claim 12; SEQ ID NO C237; 1385bp; English.
XX
XX The invention relates to nucleic acids and proteins (ADN38683-ADN40064)
CC whose expression is upregulated or downregulated in specific cancers or
CC other diseases such as angiogenic or fibrotic disorders, and to methods
CC of determining the presence or absence of a pathological cell in a
CC patient by detecting a nucleic acid at least 80% identical to those of
CC the invention or by detecting a polypeptide of the invention. The
CC invention also relates to expression vectors and host cells comprising a
CC nucleic acid of the invention; antibodies which specifically bind a
CC polypeptide of the invention; use of such antibodies for drug targeting;
CC and methods of screening for modulators of activity or expression of the
CC polypeptides and nucleic acids. The nucleic acids, polypeptides,
CC antibodies and methods are useful for diagnosing, prognosing and treating
CC cancer and other conditions such as psoriasis, leishmaniasis, heart disease,
CC atherosclerosis, inflammatory diseases, autoimmune diseases, retinal
CC neovascularization syndromes, scarring and uterine fibroids. They may
CC also be useful in wound healing and in contraception. The present
CC sequence represents a polypeptide of the invention.
XX
XX Sequence 855 AA;
SQ
Query Match 99.5%; Score 4659; DB 7; Length 855;
Best Local Similarity 99.6%; Pred. No. 9.4e-308;
Matches 852; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 MGSDBAKGKGKGPDPGAGLKYNSRHKVNGLBEGVEFLPVNNVKYKXKGPGRWVTLAA 60
DB 1 MGSDBAKGKGKGPDPGAGLKYNSRHKVNGLBEGVEFLPVNNVKYKXKGPGRWVTLAA 60
QY 61 VILGLLVLLIGIGLVVHLLQYRDYRVQKXKGYRITNENFVDAYENSSTEFYSLASKV 120
DB 61 VILGLLVLLIGIGLVVHLLQYRDYRVQKXKGYRITNENFVDAYENSSTEFYSLASKV 120
QY 121 KDALKLLYSGVPLGPGYHKSAAVTAPEGSVIATYMSFSP1POHLVEARVMAEERVVM 180
DB 121 KDALKLLYSGVPLGPGYHKSAAVTAPEGSVIATYMSFSP1POHLVEARVMAEERVVM 180
QY 181 LPPRARSLSKFSVYTSVVAFPDTSKTVORTDNCSCFGLHARGVELMFTTGPFPDSPYPA 240
DB 181 LPPRARSLSKFSVYTSVVAFPDTSKTVORTDNCSCFGLHARGVELMFTTGPFPDSPYPA 240
QY 241 HARCQWMLRGDADSVLSLTRSPFLASCDESGSLVTVNTVLSMEHALVOLCGTTPPS 300
DB 241 HARCQWMLRGDADSVLSLTRSPFLASCDESGSLVTVNTVLSMEHALVOLCGTTPPS 300
QY 301 YNLTFFSSQVNLITLTTNTERRRHGPGEATFPOLPRMSSCGGRIRKAQGFNSPYYPGHY 360
DB 301 YNLTFFSSQVNLITLTTNTERRRHGPGEATFPOLPRMSSCGGRIRKAQGFNSPYYPGHY 360
QY 361 PPNIIDCTWNIENVPNNQHVKVRFKFYYLLEBGPAGTCKPDYVEINSEKYGSGERSQFVYTS 420
DB 361 PPNIIDCTWNIENVPNNQHVKVRFKFYYLLEBGPAGTCKPDYVEINSEKYGSGERSQFVYTS 420
QY 421 NSNKITVRFHSDOSYTDGTGLAETLSYDSDPCPGQFTCTGRCIRKELACDGMADCTDH 480

DB 421 NSNKITVRFHSDOSYTDGTGLAETLSYDSDPCPGQFTCTGRCIRKELACDGMADCTDH 480
QY 481 SDELNCSGADAGHFTCKNFKCKPLFWVCSVNDGNSDQGSCEAQTFRCSNGKCLSK 540
DB 481 SDELNCSGADAGHFTCKNFKCKPLFWVCSVNDGNSDQGSCEAQTFRCSNGKCLSK 540
QY 541 SQQCGKDDCGSDSDASCEKVVVYCTKATYRCLNGLCLSKNPECDGKEDCSDSDEK 600
DB 541 SQQCGKDDCGSDSDASCEKVVVYCTKATYRCLNGLCLSKNPECDGKEDCSDSDEK 600
QY 601 DCCGGLRSFTFRQARVYVGTGTDADGEMPMQVSLHALQGHICGASLISPMVLVSAACXYID 660
DB 601 DCCGGLRSFTFRQARVYVGTGTDADGEMPMQVSLHALQGHICGASLISPMVLVSAACXYID 660
QY 661 DRGFRYSDFPQWAFPLGLHDQSORSAPGVORLTKRILISHPFNDFTFDIALLELEXP 720
DB 661 DRGFRYSDFPQWAFPLGLHDQSORSAPGVORLTKRILISHPFNDFTFDIALLELEXP 720
QY 721 ABEYSWVRPILCLPDASHVPAGRAIWTGNGHTQYGGTGAIIQKGEIRVINOJTCEML 780
DB 721 ABEYSWVRPILCLPDASHVPAGRAIWTGNGHTQYGGTGAIIQKGEIRVINOJTCEML 780
QY 781 PQQITRRMVCVGLSGVDSCCGDSGGLSVADGRIRGAGVYVSGDGCAGKNGKGYT 840
DB 781 PQQITRRMVCVGLSGVDSCCGDSGGLSVADGRIRGAGVYVSGDGCAGKNGKGYT 840
QY 841 RLPLFRDWIKENTGV 855
DB 841 RLPLFRDWIKENTGV 855
RESULT 7
ADN04754
ID ADN04754 standard; protein; 855 AA.
AC ADN04754;
XX
XX 01-JUL-2004 (first entry)
XX
XX Anti-psoriatic protein sequence #558.
XX
XX anti-psoriatic; gene therapy; psoriasis; diagnosis.
XX
XX Homo sapiens.
XX
XX WO2004028479-A2.
XX
XX 08-APR-2004.
XX
XX 25-SEP-2003; 2003MO-US030907.
XX
XX 25-SEP-2002; 2002US-0414006P.
XX
XX (GETH) GENENTECH INC.
XX
XX Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PW, Wood WJ;
PI Wu TD;
XX
XX WPI; 2004-305105/28.
DR N-PSDB; ADN04753.
XX
XX New PRO nucleic acid or polypeptide, useful for preparing a
PT pharmaceutical composition for diagnosing or treating psoriasis in a
PT mammal.
PS
PS Claim 9; SEQ ID NO 1148; 3069bp; English.
XX
XX The invention relates to novel polynucleotide and polypeptides for
CC treating psoriasis or a sequence having at least 80% identity to the
CC above sequences. The nucleic acid is useful for preparing a composition
CC for diagnosing or treating psoriasis in a mammal. This sequence
CC corresponds to one of the polypeptides of the invention.

XX Sequence 855 AA:
SQ
Query Match 99.5%; Score 4659; DB 8; Length 855;
Best Local Similarity 99.6%; Pred. No. 9,4e-308;
Matches 852; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 MGSDRAKGGGGGPGAGLKYNSRHEKVNGLSEGVLELPVNNYKVKYKQKPGKGVVLA 60
DB 1 MGSDRAKGGGGGPGAGLKYNSRHEKVNGLSEGVLELPVNNYKVKYKQKPGKGVVLA 60
QY 61 VLIGLLVLIGLGLVWHLQYRDVYQVKKNYKNTNENFVDAEYNSNSTEYSLASKV 120
DB 61 VLIGLLVLIGLGLVWHLQYRDVYQVKKNYKNTNENFVDAEYNSNSTEYSLASKV 120
QY 121 KDALKLYSGVPEFGPHKSAVTAFFSGSYIATYWSFSLPQHLVEAEKVAEERYVM 180
DB 121 KDALKLYSGVPEFGPHKSAVTAFFSGSYIATYWSFSLPQHLVEAEKVAEERYVM 180
QY 181 LPPRARSLSKSPVTVSVAFPTDSKTVOPTDQNSCSFGIHAAGVEMRFTTGPFPDSPYA 240
DB 181 LPPRARSLSKSPVTVSVAFPTDSKTVOPTDQNSCSFGIHAAGVEMRFTTGPFPDSPYA 240
QY 241 HARCQWALRGDADSVLSTFRSPDLASCDERGSDLVTYNTLSPEMHPALVOLGTYPPS 300
DB 241 HARCQWALRGDADSVLSTFRSPDLASCDERGSDLVTYNTLSPEMHPALVOLGTYPPS 300
QY 301 YNLTFFSSQNVLLTTLITNTERRHHPGFEATFQLPKMSCCGRLRKAGTNSYYRGHY 360
DB 301 YNLTFFSSQNVLLTTLITNTERRHHPGFEATFQLPKMSCCGRLRKAGTNSYYRGHY 360
QY 361 PPNIDCTMNTEVPNNQHVAFKFFYLEPVGATGPKDVEYNGEKYCGERSQFVYTS 420
DB 361 PPNIDCTMNTEVPNNQHVAFKFFYLEPVGATGPKDVEYNGEKYCGERSQFVYTS 420
QY 421 NSNKITVRFHSDQSYTGTGLAEYLSYDSDPCGQFTCRGRCIRKELRCGDADCTDH 480
DB 421 NSNKITVRFHSDQSYTGTGLAEYLSYDSDPCGQFTCRGRCIRKELRCGDADCTDH 480
QY 481 SDELNGSCDAGHQTCKNKKFCKPLFWVCDSYNDGSDNSDEGSCCPAQTFRCNSGKCLSK 540
DB 481 SDELNGSCDAGHQTCKNKKFCKPLFWVCDSYNDGSDNSDEGSCCPAQTFRCNSGKCLSK 540
QY 541 SQQNGNDGDDGSDGDEASCPKVNVTCTKHTYKCLNGCLSKGKPECDGKDCSDGSDSK 600
DB 541 SQQNGNDGDDGSDGDEASCPKVNVTCTKHTYKCLNGCLSKGKPECDGKDCSDGSDSK 600
QY 601 DCDGGLSFTQRAVVGCTDADEGEWPMQVSLHALGQHLICGASLISPMNLVSAHACYID 660
DB 601 DCDGGLSFTQRAVVGCTDADEGEWPMQVSLHALGQHLICGASLISPMNLVSAHACYID 660
QY 661 DRGRYSADPTQWTAFLGLHDSQSAAPGVQRRLKRIISHPFNDFTFDYDIALLELEKP 720
DB 661 DRGRYSADPTQWTAFLGLHDSQSAAPGVQRRLKRIISHPFNDFTFDYDIALLELEKP 720
QY 721 AEYSSMRPITCLPASHVFPAGKAIWYTGWHTYGGGALLDQKGEIRVNTQTTCELL 780
DB 721 AEYSSMRPITCLPASHVFPAGKAIWYTGWHTYGGGALLDQKGEIRVNTQTTCELL 780
QY 781 PQQITPRMVCVFGSGVDSQQDSGGLPSVEADGRIFGAGVVSWGDCAGRRKPGVYT 840
DB 781 PQQITPRMVCVFGSGVDSQQDSGGLPSVEADGRIFGAGVVSWGDCAGRRKPGVYT 840
QY 841 RLPLFRDWIKENTGV 855
DB 841 RLPLFRDWIKENTGV 855

XX 15-Apr-2004 (first entry)
DT Human NOX protein homologue Segid 420.
XX
DE human; NOX; cardiomyopathy; atherosclerosis; cancer; diabetes;
XX inflammation; autoimmune disorder; allergy; blood disorder;
KW acquired immunodeficiency syndrome; AIDS; obesity; asthma;
KW immunoglobulin (Ig)A nephropathy; cirrhosis; arthritis;
KW Alzheimer's disease; infection; str.
XX
OS Homo sapiens.
XX
PN M0200268649-A2.
XX
PD 06-SEP-2002.
XX
PE 31-JAN-2002; 2002WC-US002785.
XX
PR 31-JAN-2001; 2001US-0265395P.
PR 31-JAN-2001; 2001US-0265412P.
PR 31-JAN-2001; 2001US-0265514P.
PR 31-JAN-2001; 2001US-0265517P.
PR 02-FEB-2001; 2001US-0266406P.
PR 05-FEB-2001; 2001US-0266767P.
PR 07-FEB-2001; 2001US-0266975P.
PR 07-FEB-2001; 2001US-0267057P.
PR 08-FEB-2001; 2001US-0267459P.
PR 09-FEB-2001; 2001US-0267823P.
PR 15-FEB-2001; 2001US-0268974P.
PR 26-FEB-2001; 2001US-0271664P.
PR 27-FEB-2001; 2001US-0271839P.
PR 27-FEB-2001; 2001US-0271855P.
PR 02-MAR-2001; 2001US-0272788P.
PR 02-MAR-2001; 2001US-0273046P.
PR 14-MAR-2001; 2001US-0275925P.
PR 14-MAR-2001; 2001US-0275947P.
PR 14-MAR-2001; 2001US-0275950P.
PR 14-MAR-2001; 2001US-0275959P.
PR 15-MAR-2001; 2001US-0276448P.
PR 15-MAR-2001; 2001US-0276450P.
PR 16-MAR-2001; 2001US-0276397P.
PR 16-MAR-2001; 2001US-0276768P.
PR 20-MAR-2001; 2001US-0278652P.
PR 26-MAR-2001; 2001US-0278775P.
PR 26-MAR-2001; 2001US-0278778P.
PR 29-MAR-2001; 2001US-0279882P.
PR 29-MAR-2001; 2001US-0279884P.
PR 30-MAR-2001; 2001US-0280147P.
PR 11-APR-2001; 2001US-0282992P.
PR 11-APR-2001; 2001US-0283083P.
PR 20-APR-2001; 2001US-0285133P.
PR 23-APR-2001; 2001US-0285749P.
PR 03-MAY-2001; 2001US-0288337P.
PR 03-MAY-2001; 2001US-0288504P.
PR 29-MAY-2001; 2001US-0294047P.
PR 30-MAY-2001; 2001US-0284473P.
PR 08-JUN-2001; 2001US-0286964P.
PR 18-JUN-2001; 2001US-0298959P.
PR 19-JUN-2001; 2001US-0299324P.
PR 13-AUG-2001; 2001US-0312020P.
PR 16-AUG-2001; 2001US-0312889P.
PR 16-AUG-2001; 2001US-0312988P.
PR 21-AUG-2001; 2001US-0313390P.
PR 28-AUG-2001; 2001US-0315470P.
PR 31-AUG-2001; 2001US-0316447P.
PR 07-SEP-2001; 2001US-0318115P.
PR 07-SEP-2001; 2001US-0318118P.
PR 12-SEP-2001; 2001US-0318740P.
PR 19-SEP-2001; 2001US-0323379P.
PR 18-OCT-2001; 2001US-0330245P.
PR 18-OCT-2001; 2001US-0330308P.
PR 14-NOV-2001; 2001US-0332701P.

XX (CURA-) CURAGEN CORP.
 PA Tchernev VT, Spylek KA, Zethusen BD, Paturajan M, Shinkets RA;
 XX Li L, Gangolli EA, Radigaru M, Anderson DW, Rastelli L, Miller CE;
 PI Gelach VL, Taupier RJ, Gueev VY, Colman SD, Wolenc AR, Pena CE,
 PI Futrak K, Grosse WM, Alsobrook JP, Lepley DM, Rieger DK, Burgess CE;
 XX WPI; 2002-706998/76.
 DR New NOVX polypeptides and nucleic acids, useful for preventing or
 XX treating NOVX-associated disorders, e.g. cancer, cardiomyopathy,
 PT atherosclerosis, or diabetes, and in chromosome mapping, tissue typing or
 PT pharmacogenomics.
 PS Disclosure; SEQ ID NO 420; 1498bp; English.
 XX
 CC This invention relates to a novel nucleic acids, and encoded polypeptides
 CC thereof, which have properties related to the stimulation of biochemical
 CC or physiological responses in a cell, tissue, organ or organism.
 CC Specifically, it refers to the use of biologically active fragments for
 CC diagnostic and prognostic assays and furthermore in the treatment of
 CC diverse pathological conditions. The present invention describes novel
 CC human and murine NOVX proteins, as well as methods to modulate their
 CC expression using antisense oligos, ribozymes and peptide nucleic acids.
 CC The NOVX polypeptides, polynucleotides and antibodies are useful in
 CC treating or preventing NOVX-associated disorders, e.g. cardiomyopathy,
 CC atherosclerosis, cancer and diabetes. Furthermore, they may be used in
 CC treating or preventing diseases such as inflammation, autoimmune
 CC disorders, allergies, blood disorders, acquired immunodeficiency syndrome
 CC (AIDS), obesity, asthma, immunoglobulin (Ig)A nephropathy, cirrhosis,
 CC arthritis, Alzheimer's disease, infections, stroke, muscular dystrophy
 CC and epilepsy. Accordingly, these molecules have many activities including
 CC cytoskeletal, cardiac, anti-inflammatory, immunosuppressive, antiallergic,
 CC haemostatic, anti-HIV, antidiabetic, antiarteriosclerotic, anorectic,
 CC antiaesthetic, nephrotropic, antiarthritic, hepatotropic,
 CC neuroprotective, neurotropic, antibacterial, virucide, antiparasitic,
 CC relaxant and anticonvulsant. In addition, they are useful in screening
 CC assays to identify small molecules that modulate or inhibit, for example,
 CC neurogenesis, wound healing and angiogenesis. The nucleic acids are also
 CC used as in chromosome mapping, tissue typing, preventive medicine and
 CC pharmacogenomics. This polypeptide is a homologue of a human NOVX protein
 CC of the invention.
 CC
 XX
 XX
 SQ Sequence 855 AA;
 Query March 99.4%; Score 4655; DB 5; Length 855;
 Best Local Similarity 99.5%; Pred. No. 1.8e-307;
 Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 MGSDRARKGGGPGDFGAGLKYNSRHKVNGLEGEVFLPVNNVKKYKKGPGHVVLLAA 60
 DB 1 MGSDRARKGGGPGDFGAGLKYNSRHKVNGLEGEVFLPVNNVKKYKKGPGHVVLLAA 60
 QY 61 VLIGLLVLLIGLIGFLVHMLQYRDVAVQKXGYNRTINENFVDAYENSNSTEPLVSLASKV 120
 DB 61 VLIGLLVLLIGLIGFLVHMLQYRDVAVQKXGYNRTINENFVDAYENSNSTEPLVSLASKV 120
 QY 121 KDLAKLKYSGVPFLGYPYHKSAVTAFFSEGSVIAYVMSBPSIPQHLVEAEKVMAREEVVM 180
 DB 121 KDLAKLKYSGVPFLGYPYHKSAVTAFFSEGSVIAYVMSBPSIPQHLVEAEKVMAREEVVM 180
 QY 181 LPPRARSILKSFVYTSVVAFPDSTKVORTQDNCSCFGLHARGVLMRFTTGPFPDSYPYA 240
 DB 181 LPPRARSILKSFVYTSVVAFPDSTKVORTQDNCSCFGLHARGVLMRFTTGPFPDSYPYA 240
 QY 241 HARCQMLRGDADSVLSUTFRSFDLASCDEBGSULVTVYNTLSMEPHALVOLCGTTPPS 300
 DB 241 HARCQMLRGDADSVLSUTFRSFDLASCDEBGSULVTVYNTLSMEPHALVOLCGTTPPS 300
 QY 301 YNLTFHSSQNVLLITLLTNTERRRHGFEATFFOLPRMSSCGGRLRKXOGTFNSPYYPGHY 360
 DB 301 YNLTFHSSQNVLLITLLTNTERRRHGFEATFFOLPRMSSCGGRLRKXOGTFNSPYYPGHY 360

QY 361 PENIDCTWNIEVNNQHVKVRFFKFFYLLEBPGVAGTCPKDYVEINGEKXGERSQFVYTS 420
 DB 361 PENIDCTWNIEVNNQHVKVRFFKFFYLLEBPGVAGTCPKDYVEINGEKXGERSQFVYTS 420
 QY 421 NSNKITVRFSDDSYDTGTGLAETLSYDSSDPFGQPTCTGTGCIRKELACDGMADCTDH 480
 DB 421 NSNKITVRFSDDSYDTGTGLAETLSYDSSDPFGQPTCTGTGCIRKELACDGMADCTDH 480
 QY 481 SDELNCSGDAGHOFTCKNKKCFKPLFWVCDSVNDGSDNSDEQSCPAQTFRCSNGKLSK 540
 DB 481 SDELNCSGDAGHOFTCKNKKCFKPLFWVCDSVNDGSDNSDEQSCPAQTFRCSNGKLSK 540
 QY 541 SQQCNKGKDDCGDSDSEASCPKXNVVYTCYKTYRCLNGELCLSKNPECDKEDCSGSDSK 600
 DB 541 SQQCNKGKDDCGDSDSEASCPKXNVVYTCYKTYRCLNGELCLSKNPECDKEDCSGSDSK 600
 QY 601 DCCGGLRSFTRQARVVGCTGADGEMPMQVSLHALQGHICGASLSPNNLVSAAHCYID 660
 DB 601 DCCGGLRSFTRQARVVGCTGADGEMPMQVSLHALQGHICGASLSPNNLVSAAHCYID 660
 QY 661 DRGFRYSDFQWTAFLGILDQSORSAFVGVERLKRILSHPFNDFTFDYDIALLELEKP 720
 DB 661 DRGFRYSDFQWTAFLGILDQSORSAFVGVERLKRILSHPFNDFTFDYDIALLELEKP 720
 QY 721 AEYSSWVRPILCPDASHVFPAGKAIWYTGWGHYQYGGTGAIILOKSEIRVINOCTENLL 780
 DB 721 AEYSSWVRPILCPDASHVFPAGKAIWYTGWGHYQYGGTGAIILOKSEIRVINOCTENLL 780
 QY 781 PQQITPRMNCVGLSGGVSCQDSSGGLPSVADGRIFGAGVYSWGDCGAKNKKGVYT 840
 DB 781 PQQITPRMNCVGLSGGVSCQDSSGGLPSVADGRIFGAGVYSWGDCGAKNKKGVYT 840
 QY 841 RLPLFRDWIKENTGV 855
 DB 841 RLPLFRDWIKENTGV 855
 RESULT 9
 ADL16818
 ID ADL16818 standard; protein; 855 AA.
 XX
 AC ADL16818;
 DT 15-APR-2004 (first entry)
 XX
 DE Human NOVX protein homologue SegID 354.
 XX
 KW human; NOVX; cardiomyopathy; atherosclerosis; cancer; diabetes;
 KW inflammation; autoimmune disorder; allergy; blood disorder;
 KW acquired immunodeficiency syndrome; AIDS; obesity; asthma;
 KW immunoglobulin (Ig)A nephropathy; cirrhosis; arthritis;
 KW Alzheimer's disease; infection; str.
 XX
 OS Homo sapiens.
 XX
 PN WO200268649-A2.
 XX
 PD 06-SEP-2002.
 XX
 PF 31-JAN-2002; 2002WC-US002785.
 XX
 PR 31-JAN-2001; 2001US-0265395P.
 PR 31-JAN-2001; 2001US-0265412P.
 PR 31-JAN-2001; 2001US-0265514P.
 PR 31-JAN-2001; 2001US-0265517P.
 PR 02-FEB-2001; 2001US-0266406P.
 PR 05-FEB-2001; 2001US-0266767P.
 PR 07-FEB-2001; 2001US-0266975P.
 PR 07-FEB-2001; 2001US-0267057P.
 PR 08-FEB-2001; 2001US-0267459P.
 PR 09-FEB-2001; 2001US-0267823P.
 PR 15-FEB-2001; 2001US-0268974P.

PR 26-FEB-2001; 2001US-0271664P.
 PR 27-FEB-2001; 2001US-0271839P.
 PR 27-FEB-2001; 2001US-0271855P.
 PR 02-MAR-2001; 2001US-0272788P.
 PR 02-MAR-2001; 2001US-0273046P.
 PR 14-MAR-2001; 2001US-0275925P.
 PR 14-MAR-2001; 2001US-0275947P.
 PR 14-MAR-2001; 2001US-0275950P.
 PR 14-MAR-2001; 2001US-0275989P.
 PR 15-MAR-2001; 2001US-0276448P.
 PR 15-MAR-2001; 2001US-0276450P.
 PR 16-MAR-2001; 2001US-0276379P.
 PR 16-MAR-2001; 2001US-0276768P.
 PR 20-MAR-2001; 2001US-0278652P.
 PR 26-MAR-2001; 2001US-0278775P.
 PR 26-MAR-2001; 2001US-0278778P.
 PR 29-MAR-2001; 2001US-0279882P.
 PR 29-MAR-2001; 2001US-0279884P.
 PR 30-MAR-2001; 2001US-0280147P.
 PR 11-APR-2001; 2001US-0282992P.
 PR 11-APR-2001; 2001US-0283083P.
 PR 20-APR-2001; 2001US-0285133P.
 PR 23-APR-2001; 2001US-0285749P.
 PR 03-MAY-2001; 2001US-0288327P.
 PR 03-MAY-2001; 2001US-0288504P.
 PR 29-MAY-2001; 2001US-0294047P.
 PR 30-MAY-2001; 2001US-0294473P.
 PR 08-JUN-2001; 2001US-0296964P.
 PR 18-JUN-2001; 2001US-0298959P.
 PR 19-JUN-2001; 2001US-0299324P.
 PR 13-AUG-2001; 2001US-0312030P.
 PR 16-AUG-2001; 2001US-0312889P.
 PR 16-AUG-2001; 2001US-0312908P.
 PR 21-AUG-2001; 2001US-0313390P.
 PR 28-AUG-2001; 2001US-0315470P.
 PR 31-AUG-2001; 2001US-0316447P.
 PR 07-SEP-2001; 2001US-0318115P.
 PR 07-SEP-2001; 2001US-0318118P.
 PR 12-SEP-2001; 2001US-0318740P.
 PR 19-SEP-2001; 2001US-0323379P.
 PR 18-OCT-2001; 2001US-0330245P.
 PR 18-OCT-2001; 2001US-0330308P.
 PR 14-NOV-2001; 2001US-0332701P.
 (CURA-) CURAGEN CORP.
 XX Tchernev VT, Spytek KA, Zernusen BD, Paturajan M, Shinkets RA;
 XX Li L, Gangoli EA, Padigaru M, Anderson DM, Rastelli L, Miller CB;
 PI Gerlach VL, Taupier RJ, Gusev VY, Colman SD, Wolenc AR, Pena CE;
 PI Futak K, Grose WM, Alsbrook JF, Lepley DM, Rieger DK, Burgess CE;
 XX WPI; 2002-706998/76.
 DR New NOVX polypeptides and nucleic acids, useful for preventing or
 XX treating NOVX-associated disorders, e.g. cancer, cardiomyopathy,
 PT atherosclerosis, or diabetes, and in chromosome mapping, tissue typing or
 PT pharmacogenomics.
 XX Disclosure; SEQ ID NO 354; 1498bp; English.
 XX This invention relates to a novel nucleic acid, and encoded polypeptides
 CC thereof, which have properties related to the stimulation of biochemical
 CC or physiological responses in a cell, tissue, organ or organism.
 CC Specifically, it refers to the use of biologically active fragments for
 CC diagnostic and prognostic assays and furthermore in the treatment of
 CC diverse pathological conditions. The present invention describes novel
 CC human and murine NOVX proteins, as well as methods to modulate their
 CC expression using antisense oligos, ribozymes and peptide nucleic acids.
 CC The NOVX polypeptides, polynucleotides and antibodies are useful in
 CC treating or preventing NOVX-associated disorders, e.g. cardiomyopathy,
 CC atherosclerosis, cancer and diabetes. Furthermore, they may be used in
 CC treating or preventing diseases such as inflammation, autoimmune
 CC disorders, allergies, blood disorders, acquired immunodeficiency syndrome

CC (AIDS), obesity, asthma, immunoglobulin (Ig) A nephropathy, cirrhosis,
 CC arthritis, Alzheimer's disease, infections, stroke, muscular dystrophy
 CC and epilepsy. Accordingly, these molecules have many activities including
 CC cyostatic, cardiac, anti-inflammatory, immunosuppressive, antiallergic,
 CC hemostatic, anti-HIV, antidiabetic, antiarteriosclerotic, anorectic,
 CC antiaesthetic, nephrotoxic, antiarthritic, hepatotoxic,
 CC neuroprotective, neurotropic, antibacterial, virucide, antiparasitic,
 CC relaxant and anticonvulsant. In addition, they are useful in screening
 CC assays to identify small molecules that modulate or inhibit, for example,
 CC neurogenesis, wound healing and angiogenesis. The nucleic acids are also
 CC used as in chromosome mapping, tissue typing, preventive medicine and
 CC pharmacogenomics. This polypeptide is a homologue of a human NOVX protein
 CC of the invention.
 XX
 SQ Sequence 855 AA;
 Query Match 99.4%; Score 4655; DB 5; Length 855;
 Best Local Similarity 99.5%; Pred. No. 1.8e-307;
 Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 MGSDRAKGGGKPDPAAGLKNSRHKVNGLEGEVFLPVNVVKYKHKHGRWVLLAA 60
 DB 1 MGSDRAKGGGKPDPAAGLKNSRHKVNGLEGEVFLPVNVVKYKHKHGRWVLLAA 60
 QY VLGILLVLLIGIFLVVHLYRDVVRVQKNGYMRITNENFVDAEENSSTEFVSLASKV 120
 DB VLGILLVLLIGIFLVVHLYRDVVRVQKNGYMRITNENFVDAEENSSTEFVSLASKV 120
 QY 121 KDLAKLLYSVPLGPRYHKSAAVTAPSEGSVIAVYWSFISIPQHLVEAEARVMAEERVVM 180
 DB 121 KDLAKLLYSVPLGPRYHKSAAVTAPSEGSVIAVYWSFISIPQHLVEAEARVMAEERVVM 180
 QY 121 KDLAKLLYSVPLGPRYHKSAAVTAPSEGSVIAVYWSFISIPQHLVEAEARVMAEERVVM 180
 DB 121 KDLAKLLYSVPLGPRYHKSAAVTAPSEGSVIAVYWSFISIPQHLVEAEARVMAEERVVM 180
 QY 181 LPPPARSLKSFVYTSVVAAPFTDSKTQVORTDNCSCFELHARGVELMKEFTTGGPDSYPA 240
 DB 181 LPPPARSLKSFVYTSVVAAPFTDSKTQVORTDNCSCFELHARGVELMKEFTTGGPDSYPA 240
 QY 241 HARCQNALRGDADSVLSLTFRSFDLASCDERGSIDLTVVYVTLSPMEHALVOLGTYPPS 300
 DB 241 HARCQNALRGDADSVLSLTFRSFDLASCDERGSIDLTVVYVTLSPMEHALVOLGTYPPS 300
 QY 241 HARCQNALRGDADSVLSLTFRSFDLASCDERGSIDLTVVYVTLSPMEHALVOLGTYPPS 300
 DB 241 HARCQNALRGDADSVLSLTFRSFDLASCDERGSIDLTVVYVTLSPMEHALVOLGTYPPS 300
 QY 301 YNLFSSQNVLLITLITNTERRRHPEEATFPOLPRMSCGGRILKAQGTFFNSPYYPGHY 360
 DB 301 YNLFSSQNVLLITLITNTERRRHPEEATFPOLPRMSCGGRILKAQGTFFNSPYYPGHY 360
 QY 361 PPNIDCTWNIENVNNOHVKKRPFKFFYLLEPGVAGTCPKQVYVINGEKYGGERSQFVYTS 420
 DB 361 PPNIDCTWNIENVNNOHVKKRPFKFFYLLEPGVAGTCPKQVYVINGEKYGGERSQFVYTS 420
 QY 421 NSNKITVRPHSDSYTDTGFLAEYLSYSSDPCEGQFTCTGRCIRKELACDGMADCTDH 480
 DB 421 NSNKITVRPHSDSYTDTGFLAEYLSYSSDPCEGQFTCTGRCIRKELACDGMADCTDH 480
 QY 481 SDELNCSGADGHOFTCKNKECKPLFWYCDSDVNDGDNDSBQSCPAQTRFCSNGKCLSK 540
 DB 481 SDELNCSGADGHOFTCKNKECKPLFWYCDSDVNDGDNDSBQSCPAQTRFCSNGKCLSK 540
 QY 541 SOQNGKDDGDDSDDEASCKRVNVYVTKTKTYRCINCLCSKKNPECDGDESDGDEK 600
 DB 541 SOQNGKDDGDDSDDEASCKRVNVYVTKTKTYRCINCLCSKKNPECDGDESDGDEK 600
 QY 601 DCDGSLRSFTROARVVGADGDEMPQVSLHALGQSHLISPMNLVSAAHCYID 660
 DB 601 DCDGSLRSFTROARVVGADGDEMPQVSLHALGQSHLISPMNLVSAAHCYID 660
 QY 661 DRGFRYSDPTQWTAFLGLHDQSRASAGVQERRLKRISHPFNDFTFDYDIALLEKRP 720
 DB 661 DRGFRYSDPTQWTAFLGLHDQSRASAGVQERRLKRISHPFNDFTFDYDIALLEKRP 720
 QY 721 AEVSMWRPCLPDASVFPAGRAIWTGNGHYOGGTGALLIOKGIIRVINTOTCNLL 780
 DB 721 AEVSMWRPCLPDASVFPAGRAIWTGNGHYOGGTGALLIOKGIIRVINTOTCNLL 780
 QY 781 PQGITPRMVCVFLSGVDSGQSDGSLSSVEADGRIFGAGVWSGDCAGRNKPGVYT 840

Db 781 PQQITPRMCMCVGLSGGVDSCGDSCGPLSSVEADGRIFQAGVVSMDGCAQRNKGVT 840
Qy 841 RLPLFRDWIKENTGV 855
Db 841 RLPLFRDWIKENTGV 855

RESULT 10
AA06671
ID AAY06671 standard; protein; 855 AA.
XX AAY06671;
XX 09-NOV-1999 (first entry)
XX Tumour antigen derived gene-15 (TADG-15) protein.
XX Tumour antigen derived gene-15; TADG-15; serine protease; human;
KM breast cancer; ovary cancer; carcinoma; diagnosis.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Domain 1..54 /note= "cytoplasmic domain"
FT Domain 55..213 /note= "transmembrane domain"
FT Modified-site 109..111 /note= "Asn is N-glycosylated"
FT Region 214..447 /note= "CUB repeat"
FT Modified-site 302..304 /note= "Asn is N-glycosylated"
FT Region 453..602 /note= "ligand-binding repeat (class A motif)"
FT Region 481..483 /note= "conserved SDE motif"
FT Region 518..520 /note= "conserved SDE motif"
FT Region 554..556 /note= "conserved SDE motif"
FT Region 597..599 /note= "conserved SDE motif"
FT Region 614..615 /note= "conserved SDE motif"
FT Cleavage-site 615..855 /note= "catalytic domain"
XX
PN MO9942120-A1.
XX 26-AUG-1999.
XX 18-FEB-1999; 99WO-US003436.
XX 20-FEB-1998; 98US-00027337.
XX (UYAR-) UNIV ARKANSAS.
XX O'brien TJ, Tanimoto H;
PI WPI; 1999-527418/44.
DR N-PSDB; AAX87815.
XX
PT A new extracellular serine protease for diagnosis of neoplastic disease.
XX Claim 3; Fig 10; 71pp; English.
XX
CC The present sequence represents a novel human extracellular serine
CC protease, termed tumour antigen derived gene-15 protein (see AAY06671),
CC that is overexpressed in breast and ovarian carcinomas. The TADG-15 gene
CC (see AAX87815) can be used as a diagnostic and therapeutic target in
CC ovarian carcinoma and other carcinomas including breast, prostate, lung
CC and colon. The ligand binding domains of TADG-15 may be valuable in the

CC uptake of specific molecules into tumour cells. The invention also
CC provides: a vector that is capable of expressing DNA encoding TADG-15
CC protein; host cells selected from bacterial cells (especially Escherichia
CC coli), mammalian cells, plant cells and insect cells; and a method of
CC detecting expression of TADG-15 protein using a hybridisation probe
XX
SQ Sequence 855 AA;

Query Match 99.4%; Score 4653; DB 2; Length 855;
Best Local Similarity 99.5%; Pred. No. 2, 4e-307;
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 MGSDDRARKGGGGPKDQAGLKTNSRHKVNGLEEGVEPLPVNNVKKYKRGGRVYLA 60
Db 1 MGSDDRARKGGGGPKDQAGLKTNSRHKVNGLEEGVEPLPVNNVKKYKRGGRVYLA 60
Qy 61 VLTGLLVLTGLIGFLVWHLQYRDVRYOKVNGVMRLTNEFVAYNSNSTEVSLSKV 120
Db 61 VLTGLLVLTGLIGFLVWHLQYRDVRYOKVNGVMRLTNEFVAYNSNSTEVSLSKV 120
Qy 121 KDALKLISGVPLGPHYKESAVTAESGSLVIAVYSEFSIPQHLVEAERVNAEERVVM 180
Db 121 KDALKLISGVPLGPHYKESAVTAESGSLVIAVYSEFSIPQHLVEAERVNAEERVVM 180
Qy 181 LPPRANSLKSFVTVSVAFPTDSKTYQRTODNSCSFGLHARGVELMRFPTTGPDSFPYA 240
Db 181 LPPRANSLKSFVTVSVAFPTDSKTYQRTODNSCSFGLHARGVELMRFPTTGPDSFPYA 240
Qy 241 HARCQNALRGDADSVLSIFRSPFLASCDERGSIDLTVVYNTLSPMRHALVOLCGTYPPS 300
Db 241 HARCQNALRGDADSVLSIFRSPFLASCDERGSIDLTVVYNTLSPMRHALVOLCGTYPPS 300
Qy 301 YNLTFFHSQNVLLITLITNTERBHPGEATFQLPRMSSCGRLRAOGTFNSPYYPGHY 360
Db 301 YNLTFFHSQNVLLITLITNTERBHPGEATFQLPRMSSCGRLRAOGTFNSPYYPGHY 360
Qy 361 PNVIDCTWNIENVNNOHKVRFKFFYLLEPGVAGTCPKDYVINEBKTCGEGSOPVYVS 420
Db 361 PNVIDCTWNIENVNNOHKVRFKFFYLLEPGVAGTCPKDYVINEBKTCGEGSOPVYVS 420
Qy 421 NSNKITVRFHSDSYDIDTGLFALYLSYDSDPCPGQFTCRGTGCIKELRCMDADCTDH 480
Db 421 NSNKITVRFHSDSYDIDTGLFALYLSYDSDPCPGQFTCRGTGCIKELRCMDADCTDH 480
Qy 481 SDELNCSDAQHOFCTCKNFKCKPLFWVCDSVNDGDNSEOGSCCPAQTFRCNGKCLSK 540
Db 481 SDELNCSDAQHOFCTCKNFKCKPLFWVCDSVNDGDNSEOGSCCPAQTFRCNGKCLSK 540
Qy 541 SQCCNGKDCGDSDEASCPKXNVVCTKATYRCLNGLCLSKNPECDGKEDSDSDSK 600
Db 541 SQCCNGKDCGDSDEASCPKXNVVCTKATYRCLNGLCLSKNPECDGKEDSDSDSK 600
Qy 601 DCCGGLRSFTROARVVGTDADGEMPMOVSILALQGHICGASLISPMVLVSAACIYD 660
Db 601 DCCGGLRSFTROARVVGTDADGEMPMOVSILALQGHICGASLISPMVLVSAACIYD 660
Qy 661 DRGFRYSDFQTWATFGLHDQSORSAFGVERLKAIIISHPFNDFTFYDIALBLEKP 720
Db 661 DRGFRYSDFQTWATFGLHDQSORSAFGVERLKAIIISHPFNDFTFYDIALBLEKP 720
Qy 721 AEYSMWRPICLPDASHVPAGKAIWVTGWHQYQGTGALLIQKEIRIYNITCEML 780
Db 721 AEYSMWRPICLPDASHVPAGKAIWVTGWHQYQGTGALLIQKEIRIYNITCEML 780
Qy 781 PQQITPRMCMCVGLSGGVDSCGDSCGPLSSVEADGRIFQAGVVSMDGCAQRNKGVT 840
Db 781 PQQITPRMCMCVGLSGGVDSCGDSCGPLSSVEADGRIFQAGVVSMDGCAQRNKGVT 840
Qy 841 RLPLFRDWIKENTGV 855
Db 841 RLPLFRDWIKENTGV 855

RESULT 11
 AAB98500
 ID AAB98500 standard; protein; 855 AA.
 XX
 AC AAB98500;
 XX
 DT 03-AUG-2001 (first entry)
 XX
 DE Human TADG-15.
 XX
 KM Human; TADG-15; cytostatic; vaccine; ovarian tumour; cancer;
 KM tumour antigen-derived gene 15; extracellular serine protease.
 XX
 OS Homo sapiens.
 XX
 PN MO200129056-A1.
 XX
 PD 26-APR-2001.
 XX
 PF 20-OCT-2000; 2000MO-US029095.
 XX
 PR 20-OCT-1999; 99US-00421213.
 XX
 PA (UYAR-) UNIV ARKANSAS.
 XX
 PI O'Brien TJ, Tamimoto H;
 XX
 DR WPI; 2001-381031/40.
 DR N-PSDB; AAH23601.
 XX
 PT Novel extracellular serine protease, termed tumor antigen-derived gene 15
 PT protein overexpressed in carcinomas and DNA encoding it, for diagnosis,
 PT treatment, prevention of cancer, particularly breast, ovarian cancer.
 XX
 PS Claim 11; Fig 2; 130pp; English.
 XX
 CC The present sequence represents human tumour antigen-derived gene 15
 CC (TADG-15) protein. TADG-15 is an extracellular serine protease. It was
 CC found that TADG-15 is over-expressed in ovarian tumours. TADG-15 protein
 CC or its fragments of 9-20 residues that lack TADG-15 protease activity are
 CC useful for vaccinating an individual against TADG-15, having, suspected
 CC of having or at risk of getting cancer. Furthermore, the TADG-15 gene can
 CC be used as a diagnostic or therapeutic target in cancer
 CC
 SQ Sequence 855 AA;

Query Match 99.4%; Score 4653; DB 4; Length 855;
 Best Local Similarity 99.5%; Pred. No. 2.4e-307;
 Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 MGSRRARRGGGGPPDFGAGLKYNSRHEKVNGLSEGEVEFLPVNNVKKYKKGPRGVVTLAA 60
 DB 1 MGSRRARRGGGGPPDFGAGLKYNSRHEKVNGLSEGEVEFLPVNNVKKYKKGPRGVVTLAA 60
 QY 61 VLIGLLVLTIGLIGFLVHMLQYRDVAVOKKNGYMRITENFVDAVENSSTEFVSLASKV 120
 DB 61 VLIGLLVLTIGLIGFLVHMLQYRDVAVOKKNGYMRITENFVDAVENSSTEFVSLASKV 120
 QY 121 KDALKLYSGVPFLGPRYHKEGAVTAFSEGSVIAYYMSFSLPQHLVEABERVMAEERVVM 180
 DB 121 KDALKLYSGVPFLGPRYHKEGAVTAFSEGSVIAYYMSFSLPQHLVEABERVMAEERVVM 180
 QY 121 KDALKLYSGVPFLGPRYHKEGAVTAFSEGSVIAYYMSFSLPQHLVEABERVMAEERVVM 180
 DB 121 KDALKLYSGVPFLGPRYHKEGAVTAFSEGSVIAYYMSFSLPQHLVEABERVMAEERVVM 180
 QY 181 LPPPARSKSFVMSVVAFPPTDSKTVOQTQNSCSFGHAGVLELMRTTGGFDPSPYPA 240
 DB 181 LPPPARSKSFVMSVVAFPPTDSKTVOQTQNSCSFGHAGVLELMRTTGGFDPSPYPA 240
 QY 181 LPPPARSKSFVMSVVAFPPTDSKTVOQTQNSCSFGHAGVLELMRTTGGFDPSPYPA 240
 DB 181 LPPPARSKSFVMSVVAFPPTDSKTVOQTQNSCSFGHAGVLELMRTTGGFDPSPYPA 240
 QY 241 HARCQMALRGADSVLSTFRSPDLASCDERGSDELVTYNTLISPMPEHALVOLGCTYPPS 300
 DB 241 HARCQMALRGADSVLSTFRSPDLASCDERGSDELVTYNTLISPMPEHALVOLGCTYPPS 300
 QY 301 YNLTFHSSQNVLLITLITNTERRHGFEATFPQLPRMSSCGRLRKAQGTNSPYPGHY 360
 DB 301 YNLTFHSSQNVLLITLITNTERRHGFEATFPQLPRMSSCGRLRKAQGTNSPYPGHY 360

QY 361 PENIDCTMNIEVPNNQHVKKVFKFFYLLEPGVAGTCPKDVEINGEKYCGERSQFVVT 420
 DB 361 PENIDCTMNIEVPNNQHVKKVFKFFYLLEPGVAGTCPKDVEINGEKYCGERSQFVVT 420
 QY 421 NSNKITVRPHSDQSYDTGFLAEYLSYDSDPCPGQFTCRGRCIRKELRCGMACTDH 480
 DB 421 NSNKITVRPHSDQSYDTGFLAEYLSYDSDPCPGQFTCRGRCIRKELRCGMACTDH 480
 QY 481 SDELNCSGDAGHQTCKNKKCKPLFWVCDSVNDGDNSSDQSCSPAQTRCSNGKLSK 540
 DB 481 SDELNCSGDAGHQTCKNKKCKPLFWVCDSVNDGDNSSDQSCSPAQTRCSNGKLSK 540
 QY 541 SQQNGKDDCGDSDSEASCPVNVVCTCKHTRYCLANGCLSKNPECDGEDCSGSDDEK 600
 DB 541 SQQNGKDDCGDSDSEASCPVNVVCTCKHTRYCLANGCLSKNPECDGEDCSGSDDEK 600
 QY 601 DCDGGLRSFTRQARVVGTDADBEQWPMQVSLHALGQGHICGASLISPNMLVSAHACYID 660
 DB 601 DCDGGLRSFTRQARVVGTDADBEQWPMQVSLHALGQGHICGASLISPNMLVSAHACYID 660
 QY 661 DRGFRYSDPTQWTFELGLHDQSGRSAPGVQERRLKRIISHPFYDFTFDYDIALLELEKP 720
 DB 661 DRGFRYSDPTQWTFELGLHDQSGRSAPGVQERRLKRIISHPFYDFTFDYDIALLELEKP 720
 QY 721 AEYSMWRTPICLPASHVFPAGKAIWVTGNGHTQYGGTGALILQKGIIRVINOTTCENL 780
 DB 721 AEYSMWRTPICLPASHVFPAGKAIWVTGNGHTQYGGTGALILQKGIIRVINOTTCENL 780
 QY 781 PQQITPRMCCVFLSGVDSCQDGSGLSSVEADGRIFGAGVSWGDGCAGRNKPQVYT 840
 DB 781 PQQITPRMCCVFLSGVDSCQDGSGLSSVEADGRIFGAGVSWGDGCAGRNKPQVYT 840
 QY 841 RLPLFRDMIKENTGV 855
 DB 841 RLPLFRDMIKENTGV 855

RESULT 12
 AAE06930
 ID AAE06930 standard; protein; 855 AA.
 XX
 AC AAE06930;
 XX
 DT 16-OCT-2001 (first entry)
 XX
 DE Human membrane-type serine protease (MTSP) 1.
 XX
 KM Human; transmembrane serine protease; membrane-type serine protease;
 KM lung carcinoma; colon adenocarcinoma; ovarian carcinoma; gene therapy;
 KM matrixprase.
 XX
 OS Homo sapiens.
 XX
 PN Key location/Qualifiers
 FT Domain 615..855
 FT /label= Protease_domain
 MO200157194-A2.
 PD 09-AUG-2001.
 PD 02-FEB-2001; 2001MO-US003471.
 PF 03-FEB-2000; 2000US-0179982P.
 PR 18-FEB-2000; 2000US-0183542P.
 PR 22-JUN-2000; 2000US-0231342P.
 PR 26-JUL-2000; 2000US-0230970P.
 PR 08-SEP-2000; 2000US-0065798P.
 PR 22-SEP-2000; 2000US-0234840P.
 XX
 PA (CORV-) CORVAS INT INC.

XX Madison EL, Ong EO, Yeh J;
 XX WPI; 2001-488877/53.
 DR N-PSDB; AAD13113.
 XX Novel single chain polypeptide comprising protease domain of type-II
 PT membrane-type serine protease or its catalytically active portion useful
 PT for treating and preventing cancer and tumor.
 XX
 PS Claim 12; Page 195-197; 256pp; English.
 XX
 CC The invention relates to transmembrane serine proteases and their
 CC corresponding nucleotides and the protease domain of a type-II membrane-
 CC type serine protease (MTSP). MTSP is useful for identifying compounds
 CC that modulate or inhibit its proteolytic activity and for formulating a
 CC medicament for treating neoplastic disease. MTSP and its corresponding
 CC nucleotides are useful in preventing or treating tumors or cancers such
 CC as lung carcinoma, colon adenocarcinoma and ovarian carcinoma, in
 CC diagnostics and in hybridisation assays. MTSP is useful as a diagnostic
 CC marker for tumour development, growth and/or progression and as
 CC immunogens to generate antibodies that specifically bind to it. MTSP DNA
 CC is useful in a yeast two-hybrid system and in gene therapy. The present
 CC sequence is human MTSP1 protein (also called matrilysin).
 XX
 SQ Sequence 855 AA;
 Query Match 99.4%; Score 4653; DB 4; Length 855;
 Best Local Similarity 99.5%; Pred. No. 2,4e-307;
 Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 MGSDBARKGGGGPKFGAGLKYNSRHEKVNGLBEVGFLLPVNNVKKYKIGPGRVVLAA 60
 DB 1 MGSDBARKGGGGPKFGAGLKYNSRHEKVNGLBEVGFLLPVNNVKKYKIGPGRVVLAA 60
 QY 61 VLGLLLVLLIGLTVLHLYQYDVRVQKXNKNYRINENPVDAYENSNTSEFVSLASKY 120
 DB 61 VLGLLLVLLIGLTVLHLYQYDVRVQKXNKNYRINENPVDAYENSNTSEFVSLASKY 120
 QY 121 KDALKLILYSGVPFPGPHKESAVTAFSEGSVIAYYSEFSLPQHLVVEARVMAEERVM 180
 DB 121 KDALKLILYSGVPFPGPHKESAVTAFSEGSVIAYYSEFSLPQHLVVEARVMAEERVM 180
 QY 181 LPPRARSLSKFPVVTSSVVAFPDTSKTVQRTDNGSCFGLHARGVLMFTTPGFPDSPA 240
 DB 181 LPPRARSLSKFPVVTSSVVAFPDTSKTVQRTDNGSCFGLHARGVLMFTTPGFPDSPA 240
 QY 241 HARCQWALRGDADSVSLTFRSPDLASCDRGSLLVYVNTLSMEPHALVOLCGTYPSP 300
 DB 241 HARCQWALRGDADSVSLTFRSPDLASCDRGSLLVYVNTLSMEPHALVOLCGTYPSP 300
 QY 301 YNLTFRSSQVLLTLTLTNTERRHPGEATFPQLPRMSSCGRLRKQGFNSPYPYGHY 360
 DB 301 YNLTFRSSQVLLTLTLTNTERRHPGEATFPQLPRMSSCGRLRKQGFNSPYPYGHY 360
 QY 361 PENIDCTWNIETVNNQHVKRFKFFYLLEPGVPAGTCKDYVEINGEKYCGERSQFVVT 420
 DB 361 PENIDCTWNIETVNNQHVKRFKFFYLLEPGVPAGTCKDYVEINGEKYCGERSQFVVT 420
 QY 421 NSKKITRFRHSDDSYDTGTGLAEYLSVSDSDPCGQGTCTGRTIRKELACDGMADCTDH 480
 DB 421 NSKKITRFRHSDDSYDTGTGLAEYLSVSDSDPCGQGTCTGRTIRKELACDGMADCTDH 480
 QY 481 SDEINSCDAGHQPCTKNKCKPLFWVCDSVNDGSDSDSGSCPAQTRFCSNGKCLSK 540
 DB 481 SDEINSCDAGHQPCTKNKCKPLFWVCDSVNDGSDSDSGSCPAQTRFCSNGKCLSK 540
 QY 541 SQQCNKGKDCGSDGDEASCPKVVNVVTKTKHYRCLNGLCLSKNGPACDGEDSGDSDG 600
 DB 541 SQQCNKGKDCGSDGDEASCPKVVNVVTKTKHYRCLNGLCLSKNGPACDGEDSGDSDG 600
 QY 601 DCCGGLRSFTROARVVGCTADBEEMPMQVSHALGQGHICGASLISPMWLVSAAHCYTD 660
 DB 601 DCCGGLRSFTROARVVGCTADBEEMPMQVSHALGQGHICGASLISPMWLVSAAHCYTD 660

DB 601 DCCGGLRSFTROARVVGCTADBEEMPMQVSHALGQGHICGASLISPMWLVSAAHCYTD 660
 QY 661 DRGFRYSDDPTQMTAFGLHDQSORSAPGVQERLKRILISHPFNDFTFDYDIALLELEKP 720
 DB 661 DRGFRYSDDPTQMTAFGLHDQSORSAPGVQERLKRILISHPFNDFTFDYDIALLELEKP 720
 QY 721 AEYSWVRPILCPDASHVPFAGKAIWVTGHTQYGGTGLILQKBEIRVINTTCBNLL 780
 DB 721 AEYSWVRPILCPDASHVPFAGKAIWVTGHTQYGGTGLILQKBEIRVINTTCBNLL 780
 QY 781 PQQITRRMVCVGLSGVSCGDSGGLSSVSEADRIKAGVAVSWGDCAGRNKCVYT 840
 DB 781 PQQITRRMVCVGLSGVSCGDSGGLSSVSEADRIKAGVAVSWGDCAGRNKCVYT 840
 QY 841 RLPLFRDWIKENTGV 855
 DB 841 RLPLFRDWIKENTGV 855
 RESULT 13
 AAO22929
 ID AAO22929 standard; protein; 855 AA.
 AC AAO22929;
 XX 12-DEC-2002 (first entry)
 DE Type II transmembrane serine protease 1 protein SEQ ID No 2.
 XX
 KW Cytostatic; type-II membrane-type serine protease 7; MTSP7; malignancy;
 KW neoplastic disease; pre-malignant lesion; tumour; zymogen form; cancer;
 KW malignant; enzyme.
 XX Homo sapiens.
 XX NO200272786-A2.
 XX 19-SEP-2002.
 PF 13-MAR-2002; 2002MO-US007903.
 XX 13-MAR-2001; 2001US-0275592P.
 PR (CORV-) CORVAS INT INC.
 PA Madison EL, Ong EO;
 XX WPI; 2002-732827/79.
 DR N-PSDB; AAL53444.
 XX
 PT New transmembrane serine protease 7 (MTSP7) polypeptide for diagnosing
 PT neoplastic diseases, monitoring tumor progress or therapeutic
 PT effectiveness, or identifying MTSP7 modulators for treating tumors or
 PT cancers.
 XX
 PS Disclosure; Page 172-174; 184pp; English.
 XX
 CC The invention relates to a purified single or two-chain polypeptide,
 CC which comprises the protease domain of a type-II membrane-type serine
 CC protease 7 (MTSP7) or its catalytically active portion. The polypeptide
 CC comprising MTSP7 is useful for detecting or diagnosing a neoplastic
 CC disease, a pre-malignant lesion, a malignancy or other pathological
 CC condition in a subject. This polypeptide is also useful for monitoring
 CC tumor (e.g. tumor of the breast, cervix, prostate, lung, ovary or
 CC colon) progress and/or therapeutic effectiveness. The inhibitor of the
 CC polypeptide or the inhibitor of the polypeptide's zymogen form is useful
 CC for treating or preventing a neoplastic disease, or tumor initiation,
 CC growth or progression, or a (pre-)malignant condition. The polypeptide or
 CC polynucleotide is also useful for identifying modulators of MTSP7, which
 CC may be used to treat cancers or tumors. This sequence represents a
 CC protein of the type-II membrane-type serine protease 1 relating to the
 CC invention
 XX

Sequence 855 AA;
Query Match 99.4%; Score 4653; DB 5; Length 855;
Best Local Similarity 99.5%; Pred. No. 2,4e-307;
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 MGSDRAKGGGGPPDFGAGLKYNSRHEKYNGLSEGEVEFLPVNNYKVKYKPGHVVVLA 60
D 1 MGSDRAKGGGGPPDFGAGLKYNSRHEKYNGLSEGEVEFLPVNNYKVKYKPGHVVVLA 60
QY 61 VLIGLLVLIGLIGFLVHLYQYRDYRVQKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKX 120
D 61 VLIGLLVLIGLIGFLVHLYQYRDYRVQKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKX 120
QY 121 KDALKLYSGVPFLGPHYKESAVTAFSEGSVIATYMSFSI.PQHLVEAEKRVMAEERVVM 180
D 121 KDALKLYSGVPFLGPHYKESAVTAFSEGSVIATYMSFSI.PQHLVEAEKRVMAEERVVM 180
QY 181 LPPARSLKSFVYVSVAFFPDSKTQRTQDSCSFGIHAHAGVELMREPTTGPDPSPYPA 240
D 181 LPPARSLKSFVYVSVAFFPDSKTQRTQDSCSFGIHAHAGVELMREPTTGPDPSPYPA 240
QY 241 HARCQMALRGDADSVLSTFRSPDLASCDERGSBLVTYNTLSPEMHPALVQLGTYPPS 300
D 241 HARCQMALRGDADSVLSTFRSPDLASCDERGSBLVTYNTLSPEMHPALVQLGTYPPS 300
QY 301 YNLTFHSSQNVLLTLITNTERRHGPFATFPOLPRMSSCGGRIRKAKOGTNSPYRGHY 360
D 301 YNLTFHSSQNVLLTLITNTERRHGPFATFPOLPRMSSCGGRIRKAKOGTNSPYRGHY 360
QY 361 PPNIDCTWNIENVNNQHKVAFKFFYLLEPGVATGCKDVEYNGEKYCGERQOFVYTS 420
D 361 PPNIDCTWNIENVNNQHKVAFKFFYLLEPGVATGCKDVEYNGEKYCGERQOFVYTS 420
QY 421 NSNKITVRFHSDQSYTDGFLAEYLSYDSDPCQOFTCRGTGRCIRKELACDGMADCTDH 480
D 421 NSNKITVRFHSDQSYTDGFLAEYLSYDSDPCQOFTCRGTGRCIRKELACDGMADCTDH 480
QY 481 SDELNCGDAGHQTCKRKFCKPLFWCDSDVNDGDNDSDEGSCCPAOTPRCSNGKCLSK 540
D 481 SDELNCGDAGHQTCKRKFCKPLFWCDSDVNDGDNDSDEGSCCPAOTPRCSNGKCLSK 540
QY 541 SQQNGKDDCGDSDDEASCPKVNVTCTKHYRCIANGCLSKNGPECDGKEDCSGSDSK 600
D 541 SQQNGKDDCGDSDDEASCPKVNVTCTKHYRCIANGCLSKNGPECDGKEDCSGSDSK 600
QY 601 DCDGGLRSFTRQARVVGTDADBEQWQVSLHALGQGHICGASLISPNNLVSAHACYID 660
D 601 DCDGGLRSFTRQARVVGTDADBEQWQVSLHALGQGHICGASLISPNNLVSAHACYID 660
QY 661 DRGRYSDPTQWTFGLIHDSQSRAPGVQRRRLKRIISHPFNDFTFDYIALLELKP 720
D 661 DRGRYSDPTQWTFGLIHDSQSRAPGVQRRRLKRIISHPFNDFTFDYIALLELKP 720
QY 721 AEYSWVRPILCPASHVFPAGKAIWVGWHTQGGIALLQGBELRVNQTCEML 780
D 721 AEYSWVRPILCPASHVFPAGKAIWVGWHTQGGIALLQGBELRVNQTCEML 780
QY 781 PQQITPRMVCVFLSGGVDSQCGDGGFLSSVEADGRIFGAGVSWGDGAGRRKPGVYT 840
D 781 PQQITPRMVCVFLSGGVDSQCGDGGFLSSVEADGRIFGAGVSWGDGAGRRKPGVYT 840
QY 841 RLPLFRDWMKENTGV 855
D 841 RLPLFRDWMKENTGV 855

RESULT 14
AD116816
ID AD116816 standard; protein, 855 AA.
XX AC AD116816;
XX

DT 15-APR-2004 (first entry)
XX
DE Human NOX protein homologue Segid 352.
XX human; NOX; cardiomyopathy; atherosclerosis; cancer; diabetes;
KW inflammation; autoimmune disorder; allergy; blood disorder;
KW acquired immunodeficiency syndrome; AIDS; obesity; asthma;
KW immunoglobulin (Ig) A nephropathy; cirrhosis; arthritis;
KW Alzheimer's disease; infection; str.
XX
OS Homo sapiens.
PN W020268649-A2.
XX
PD 06-SEP-2002.
XX
PF 31-JAN-2002; 2002WC-US002785.
XX
PR 31-JAN-2001; 2001US-0265395P.
PR 31-JAN-2001; 2001US-0265412P.
PR 31-JAN-2001; 2001US-0265514P.
PR 31-JAN-2001; 2001US-0265517P.
PR 02-FEB-2001; 2001US-0266406P.
PR 05-FEB-2001; 2001US-0266767P.
PR 07-FEB-2001; 2001US-0266975P.
PR 07-FEB-2001; 2001US-0267057P.
PR 08-FEB-2001; 2001US-0267459P.
PR 09-FEB-2001; 2001US-0267823P.
PR 15-FEB-2001; 2001US-0268974P.
PR 26-FEB-2001; 2001US-0271664P.
PR 27-FEB-2001; 2001US-0271839P.
PR 27-FEB-2001; 2001US-0271855P.
PR 02-MAR-2001; 2001US-0272788P.
PR 02-MAR-2001; 2001US-0273046P.
PR 14-MAR-2001; 2001US-0275925P.
PR 14-MAR-2001; 2001US-0275947P.
PR 14-MAR-2001; 2001US-0275950P.
PR 14-MAR-2001; 2001US-0275989P.
PR 15-MAR-2001; 2001US-0276448P.
PR 15-MAR-2001; 2001US-0276450P.
PR 16-MAR-2001; 2001US-0276397P.
PR 16-MAR-2001; 2001US-0276788P.
PR 20-MAR-2001; 2001US-0278652P.
PR 26-MAR-2001; 2001US-0278775P.
PR 26-MAR-2001; 2001US-0278778P.
PR 29-MAR-2001; 2001US-0279882P.
PR 29-MAR-2001; 2001US-0279884P.
PR 30-MAR-2001; 2001US-0280147P.
PR 11-APR-2001; 2001US-0282992P.
PR 11-APR-2001; 2001US-0283083P.
PR 20-APR-2001; 2001US-0285133P.
PR 23-APR-2001; 2001US-0285749P.
PR 03-MAY-2001; 2001US-0288377P.
PR 03-MAY-2001; 2001US-0288504P.
PR 29-MAY-2001; 2001US-0294047P.
PR 30-MAY-2001; 2001US-0294473P.
PR 08-JUN-2001; 2001US-0296564P.
PR 18-JUN-2001; 2001US-0298959P.
PR 19-JUN-2001; 2001US-0299349P.
PR 13-AUG-2001; 2001US-0312020P.
PR 16-AUG-2001; 2001US-0312889P.
PR 16-AUG-2001; 2001US-0312988P.
PR 21-AUG-2001; 2001US-0313390P.
PR 28-AUG-2001; 2001US-0315470P.
PR 31-AUG-2001; 2001US-0316447P.
PR 07-SEP-2001; 2001US-0318115P.
PR 07-SEP-2001; 2001US-0318118P.
PR 12-SEP-2001; 2001US-0318740P.
PR 19-SEP-2001; 2001US-0323379P.
PR 18-OCT-2001; 2001US-0330245P.
PR 18-OCT-2001; 2001US-0330308P.
PR 14-NOV-2001; 2001US-0332701P.
XX

PA (CURA-) CURAGEN CORP.
XX Tchernev VT, Spytek KA, Zernusen BD, Paturajan M, Shinkens RA;
PI Li L, Ganggoli EA, Padigan M, Anderson SD, Rastelli L, Miller CE;
PI Gerlach VL, Taupier RO, Gueev VY, Colman DW, Wolenc AR, Pena CEA,
PI Futrak K, Grosse WM, Alsobrook JP, Lepley DM, Rieger DK, Burgess CE;
XX WPI; 2002-706998/76.
DR
XX New NOVX polypeptides and nucleic acids, useful for preventing or
PT treating NOVX-associated disorders, e.g. cancer, cardiomyopathy,
PT atherosclerosis, or diabetes, and in chromosome mapping, tissue typing or
PT pharmacogenomics.
PS Disclosure; SEQ ID NO 352; 1498bp; English.
XX
XX This invention relates to a novel nucleic acids, and encoded polypeptides
CC thereof, which have properties related to the stimulation of biochemical
CC or physiological responses in a cell, tissue, organ or organism.
CC Specifically, it refers to the use of biologically active fragments for
CC diagnostic and prognostic assays and furthermore in the treatment of
CC diverse pathological conditions. The present invention describes novel
CC human and murine NOVX proteins, as well as methods to modulate their
CC expression using antisense oligos, ribozymes and peptide nucleic acids.
CC The NOVX polypeptides, polynucleotides and antibodies are useful in
CC treating or preventing NOVX-associated disorders, e.g. cardiomyopathy,
CC atherosclerosis, cancer and diabetes. Furthermore, they may be used in
CC treating or preventing diseases such as inflammation, autoimmune
CC disorders, allergies, blood disorders, acquired immunodeficiency syndrome
CC (AIDS), obesity, asthma, immunoglobulin (Ig)A nephropathy, cirrhosis,
CC arthritis, Alzheimer's disease, infections, stroke, muscular dystrophy
CC and epilepsy. Accordingly, these molecules have many activities including
CC cyostatic, cardiac, antiinflammatory, immunosuppressive, antiallergic,
CC haemostatic, anti-HIV, antidiabetic, antiatherosclerotic, anorectic,
CC antiaesthetic, nephrotoxic, antiarthritic, hepatotropic,
CC neuroprotective, neurotropic, antibacterial, virucide, antiparasitic,
CC relaxant and anticonvulsant. In addition, they are useful in screening
CC assays to identify small molecules that modulate or inhibit, for example,
CC neurogenesis, wound healing and angiogenesis. The nucleic acids are also
CC used as in chromosome mapping, tissue typing, preventive medicine and
CC pharmacogenomics. This polypeptide is a homologue of a human NOVX protein
CC of the invention.
XX
SQ Sequence 855 AA;
Query Match 99.4%; Score 4653; DB 5; Length 855;
Best Local Similarity 99.5%; Pred. No. 2.4e-307;
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 MGSDRARKGGGPGKDFGAGLKYNSRHKVNGLEBGEVFLPVNNVKKYKKGPGRWVLA 60
DB 1 MGSDRARKGGGPGKDFGAGLKYNSRHKVNGLEBGEVFLPVNNVKKYKKGPGRWVLA 60
QY 61 VLIGLLVLLIGLIGLVNHLQYRDYRVOKVKNKYRTNENFVDAEYNSNSTEPIASLV 120
DB 61 VLIGLLVLLIGLIGLVNHLQYRDYRVOKVKNKYRTNENFVDAEYNSNSTEPIASLV 120
QY 121 KDAKLKLYSGVPLFGPYHKSAVTAFASEGVIATYMESEFSIPHLVEAEAEVMAEERVM 180
DB 121 KDAKLKLYSGVPLFGPYHKSAVTAFASEGVIATYMESEFSIPHLVEAEAEVMAEERVM 180
QY 181 LPPRARSILKSFVNTSVVAFPTDSKTIVORTDNCSCFGLHARGVLEMTTGPFPDSYPA 240
DB 181 LPPRARSILKSFVNTSVVAFPTDSKTIVORTDNCSCFGLHARGVLEMTTGPFPDSYPA 240
QY 241 HARQVRLKGDADSVLSITPSSFDLACDERGSLVYVNTLSMEBHALVOLCGTTPPS 300
DB 241 HARQVRLKGDADSVLSITPSSFDLACDERGSLVYVNTLSMEBHALVOLCGTTPPS 300
QY 301 YNLFFHSSQNVLTLLTITNTERRHGPPEATPFOLPRMSSCGGRLRKAQGFNSPYRGHY 360
DB 301 YNLFFHSSQNVLTLLTITNTERRHGPPEATPFOLPRMSSCGGRLRKAQGFNSPYRGHY 360

QY 361 PENIDCTWNIENVNNOHVKVRFEKPYLLBPGVAGTCPKDYVINEKTKYCGERSQFVNTS 420
DB 361 PENIDCTWNIENVNNOHVKVSFFKPYLLBPGVAGTCPKDYVINEKTKYCGERSQFVNTS 420
QY 421 NSKITYRFRHSDDSYDTGTGLAEYLSYSSDPGQPTCTGTGCTIKELRCDDWACTTH 480
DB 421 NSKITYRFRHSDDSYDTGTGLAEYLSYSSDPGQPTCTGTGCTIKELRCDDWACTTH 480
QY 481 SDELINCSQDAGHOFTCKNFKCKPLFWVCDSVNDGCGNSDEQSCPAQTFRCNGKLSK 540
DB 481 SDELINCSQDAGHOFTCKNFKCKPLFWVCDSVNDGCGNSDEQSCPAQTFRCNGKLSK 540
QY 541 SQQCNKGKDDCGSDSDASCPKVVVVTCTKHTRYCLNGLCLSKNPECDGKEDCSGSDX 600
DB 541 SQQCNKGKDDCGSDSDASCPKVVVVTCTKHTRYCLNGLCLSKNPECDGKEDCSGSDX 600
QY 601 DCCCGLRSPTRQARVVGCTDADGEMPMQVSLHALGQHCASLISPMWLSAAHCYTD 660
DB 601 DCCCGLRSPTRQARVVGCTDADGEMPMQVSLHALGQHCASLISPMWLSAAHCYTD 660
QY 661 DRGFRYSDFPTQWTAFLGLHDQSORSAPOVERLKRILISHPFNDFTPDYDIALLEK 720
DB 661 DRGFRYSDFPTQWTAFLGLHDQSORSAPOVERLKRILISHPFNDFTPDYDIALLEK 720
QY 721 AEYSSWVRPICLPDASHVPAGKAIWVTGSHGTQYGTGALLIQKEIRVINGTCENTL 780
DB 721 AEYSSWVRPICLPDASHVPAGKAIWVTGSHGTQYGTGALLIQKEIRVINGTCENTL 780
QY 781 PQGTRRMVCVGLSGGVNDSQDGGPLSSVADRIIGAGVSVGDCAGNKRKYVT 840
DB 781 PQGTRRMVCVGLSGGVNDSQDGGPLSSVADRIIGAGVSVGDCAGNKRKYVT 840
QY 841 RLPLFRDMIKENTGV 855
DB 841 RLPLFRDMIKENTGV 855
RESULT 15
AD116882
ID AD116882 standard; protein; 855 AA.
XX
AC AD116882;
DT 15-APR-2004 (first entry)
XX
DE Human NOVX protein homologue SegID 418.
XX
KW human; NOVX; cardiomyopathy; atherosclerosis; cancer; diabetes;
KW inflammation; autoimmune disorder; allergy; blood disorder;
KW acquired immunodeficiency syndrome; AIDS; obesity; asthma;
KW immunoglobulin (Ig)A nephropathy; cirrhosis; arthritis;
KW Alzheimer's disease; infection; str.
XX
OS Homo sapiens.
PN
PN WO200268649-A2.
XX
PD 06-SEP-2002.
PF 31-JAN-2002; 2002WO-US002785.
XX
PR 31-JAN-2001; 2001US-0265395P.
PR 31-JAN-2001; 2001US-0265412P.
PR 31-JAN-2001; 2001US-0265514P.
PR 31-JAN-2001; 2001US-0265517P.
PR 02-FEB-2001; 2001US-0266406P.
PR 05-FEB-2001; 2001US-0266767P.
PR 07-FEB-2001; 2001US-0266975P.
PR 07-FEB-2001; 2001US-0267057P.
PR 08-FEB-2001; 2001US-0267459P.
PR 09-FEB-2001; 2001US-0267823P.
PR 15-FEB-2001; 2001US-0268974P.
PR 26-FEB-2001; 2001US-0271664P.

PR 27-FEB-2001; 2001US-0271839P.
PR 27-FEB-2001; 2001US-0271855P.
PR 02-MAR-2001; 2001US-0272788P.
PR 02-MAR-2001; 2001US-0273046P.
PR 14-MAR-2001; 2001US-0275925P.
PR 14-MAR-2001; 2001US-0275925P.
PR 14-MAR-2001; 2001US-0275950P.
PR 14-MAR-2001; 2001US-0275989P.
PR 15-MAR-2001; 2001US-0276448P.
PR 15-MAR-2001; 2001US-0276450P.
PR 16-MAR-2001; 2001US-0276397P.
PR 16-MAR-2001; 2001US-0276768P.
PR 20-MAR-2001; 2001US-0278652P.
PR 26-MAR-2001; 2001US-0278775P.
PR 26-MAR-2001; 2001US-0278778P.
PR 29-MAR-2001; 2001US-0279882P.
PR 29-MAR-2001; 2001US-0279884P.
PR 30-MAR-2001; 2001US-0280147P.
PR 11-APR-2001; 2001US-0282992P.
PR 11-APR-2001; 2001US-0283083P.
PR 20-APR-2001; 2001US-0285133P.
PR 23-APR-2001; 2001US-0285749P.
PR 03-MAY-2001; 2001US-0288327P.
PR 29-MAY-2001; 2001US-0288504P.
PR 30-MAY-2001; 2001US-0294473P.
PR 08-JUN-2001; 2001US-0296964P.
PR 18-JUN-2001; 2001US-0298959P.
PR 19-JUN-2001; 2001US-0299324P.
PR 13-AUG-2001; 2001US-0312020P.
PR 16-AUG-2001; 2001US-0312889P.
PR 16-AUG-2001; 2001US-0312908P.
PR 21-AUG-2001; 2001US-0313398P.
PR 28-AUG-2001; 2001US-0315470P.
PR 31-AUG-2001; 2001US-0316447P.
PR 07-SEP-2001; 2001US-0318115P.
PR 07-SEP-2001; 2001US-0318115P.
PR 12-SEP-2001; 2001US-0318740P.
PR 19-SEP-2001; 2001US-0323379P.
PR 18-OCT-2001; 2001US-0330245P.
PR 18-OCT-2001; 2001US-0330308P.
PR 14-NOV-2001; 2001US-0332701P.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Tchernev VT, Spytek KA, Zernusen BD, Patturajan M, Shimkets RA;
PI Li L, Gangoli EA, Padigaru M, Anderson DW, Rastelli L, Miller CB;
PI Gerlach VT, Taupier RJ, Gusev VY, Colman SD, Wolenc AR, Pena CE;
PI Putak K, Grose WM, Alsebrook JP, Lepley DM, Rieger DK, Burgess CB;
XX
XX WPI; 2002-706998/76.
XX
XX New NOXV polypeptides and nucleic acids, useful for preventing or
PT treating NOXV-associated disorders, e.g. cancer, cardiomyopathy,
PT atherosclerosis, or diabetes, and in chromosome mapping, tissue typing or
PT pharmacogenomics.
XX
XX Disclosure; SEQ ID NO 418; 1498bp; English.
XX
XX This invention relates to a novel nucleic acid, and encoded polypeptides
CC thereof, which have properties related to the stimulation of biochemical
CC or physiological responses in a cell, tissue, organ or organism.
CC Specifically, it refers to the use of biologically active fragments for
CC diagnostic and prognostic assays and furthermore in the treatment of
CC diverse pathological conditions. The present invention describes novel
CC human and murine NOXV proteins, as well as methods to modulate their
CC expression using antisense oligos, ribozymes and peptide nucleic acids.
CC The NOXV polypeptides, polynucleotides and antibodies are useful in
CC treating or preventing NOXV-associated disorders, e.g. cardiomyopathy,
CC atherosclerosis, cancer and diabetes. Furthermore, they may be used in
CC treating or preventing diseases such as inflammation, autoimmune
CC disorders, allergies, blood disorders, acquired immunodeficiency syndrome
CC (AIDS), obesity, asthma, immunoglobulin (IgA) nephropathy, cirrhosis,

CC arthritis, Alzheimer's disease, infections, stroke, muscular dystrophy
CC and epilepsy. Accordingly, these molecules have many activities including
CC cytostatic, cardiant, antidiabetic, antitumor, immunosuppressive, antiallergic,
CC haemostatic, anti-HIV, antidiabetic, antitumor, antitoxic, anorectic,
CC antiaesthetic, nephrotoxic, antihypertensive, hepatotoxic,
CC neuroprotective, neurotropic, antibacterial, virucide, antiparasitic,
CC relaxant and anticonvulsant. In addition, they are useful in screening
CC assays to identify small molecules that modulate or inhibit, for example,
CC neurogenesis, wound healing and angiogenesis. The nucleic acids are also
CC used as in chromosome mapping, tissue typing, preventive medicine and
CC pharmacogenomics. This polypeptide is a homologue of a human NOXV protein
CC of the invention.
XX
XX Sequence 855 AA;
SQ
Query Match 99.4%; Score 4653; DB 5; Length 855;
Best Local Similarity 99.5%; Pred. No. 2,4e-307;
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 MGSDRARKGGGPRDFAGLKYNSRHEKVGLEGEVFLPVNNVKKYKKGPRMVLAA 60
DB 1 MGSDRARKGGGPRDFAGLKYNSRHEKVGLEGEVFLPVNNVKKYKKGPRMVLAA 60
QY VLGILLVLLIGIGFLVWHLQYRDVRYQKXNGYMRITNENFVDAYENSSTEFVSLASKV 120
DB VLGILLVLLIGIGFLVWHLQYRDVRYQKXNGYMRITNENFVDAYENSSTEFVSLASKV 120
QY 121 KDALKLYSGVPLGIPYHKSAYTAPEGSVIAIYMSSEFIPOHLYBEARVAEEVVM 180
DB 121 KDALKLYSGVPLGIPYHKSAYTAPEGSVIAIYMSSEFIPOHLYBEARVAEEVVM 180
QY 121 LPPRARSLSKFPVTSVAFPTDSKTQVORTODNSCSFGLHARGVBLNFTTGGPDSRYPA 240
DB 181 LPPRARSLSKFPVTSVAFPTDSKTQVORTODNSCSFGLHARGVBLNFTTGGPDSRYPA 240
QY 241 HARCQWALRGDADSVLSLTPRSFPLASCDERGSGLVYVNTLSPMERHALVOLCGTYPSS 300
DB 241 HARCQWALRGDADSVLSLTPRSFPLASCDERGSGLVYVNTLSPMERHALVOLCGTYPSS 300
QY 301 YNLTFHSSQVNLITLITNTERRHPGFEATFPLPRMSCGGRLLRKAQGTFFSPYPGHY 360
DB 301 YNLTFHSSQVNLITLITNTERRHPGFEATFPLPRMSCGGRLLRKAQGTFFSPYPGHY 360
QY 361 PPINDCTWNIENPVNNQVKKVFPKFFYLBERPVAGTCRKYVINGKYGESRQFVYTS 420
DB 361 PPINDCTWNIENPVNNQVKKVFPKFFYLBERPVAGTCRKYVINGKYGESRQFVYTS 420
QY 421 NSNKITVRFHSDSYTDTGFLAAYLSYSSDPGQFTCRGRCIRKELACDGMADCTDH 480
DB 421 NSNKITVRFHSDSYTDTGFLAAYLSYSSDPGQFTCRGRCIRKELACDGMADCTDH 480
QY 481 SDELINSCDAAGHOFCKNFKCPFLFVYCDSDVNDGSDSDEQSCSPAQTRFCSNGKLSK 540
DB 481 SDELINSCDAAGHOFCKNFKCPFLFVYCDSDVNDGSDSDEQSCSPAQTRFCSNGKLSK 540
QY 541 SQQNGKDDGSDSDASCPKVVVYCTKRTYCLNGSLCSKSNPBDGKEDSDGSDSK 600
DB 541 SQQNGKDDGSDSDASCPKVVVYCTKRTYCLNGSLCSKSNPBDGKEDSDGSDSK 600
QY 601 DCDGSLSFTRQARVVGTDADGEMPMQVSLHALGSHICGSLISPMNLVSAHACYID 660
DB 601 DCDGSLSFTRQARVVGTDADGEMPMQVSLHALGSHICGSLISPMNLVSAHACYID 660
QY 661 DRGRYSDFPQWTAFLGLHDQSORSAPOVERLRKIIISHPFNDFPFDIALLEKXP 720
DB 661 DRGRYSDFPQWTAFLGLHDQSORSAPOVERLRKIIISHPFNDFPFDIALLEKXP 720
QY 721 AYSWMPICLPDASHVPAGKAIWVGSHQYSGTGLILQKEIRVINTTCENIL 780
DB 721 AYSWMPICLPDASHVPAGKAIWVGSHQYSGTGLILQKEIRVINTTCENIL 780
QY 781 PQQITPRMVCVGLSGVDSCDSDGGLSVADAGTIFGAGVYWGDCAGRNKQGVYT 840
DB 781 PQQITPRMVCVGLSGVDSCDSDGGLSVADAGTIFGAGVYWGDCAGRNKQGVYT 840


```

QY 181 LPPRABSLKSVTVSVVAFPTDSKTVOQTODNSCSFGIHAAGVELMRPTTGPDPSPVA 240
DB 181 LPPRABSLKSVTVSVVAFPTDSKTVOQTODNSCSFGIHAAGVELMRPTTGPDPSPVA 240
QY 241 HARCQMLRGDADSVLSLTFRSPDLASCDERGSDLVTVYNTLSPEMPEHALVOLCGTYPSP 300
DB 241 HARCQMLRGDADSVLSLTFRSPDLASCDERGSDLVTVYNTLSPEMPEHALVOLCGTYPSP 300
QY 301 YNLTFFSSQNVLLITLITNTERRHPGFEATFQOLPRMSSCGGLRKAQGTFFNSPYYPGHY 360
DB 301 YNLTFFSSQNVLLITLITNTERRHPGFEATFQOLPRMSSCGGLRKAQGTFFNSPYYPGHY 360
QY 361 PPNIDCTWNIENVNNOHVKKRFKFFYLLEPGVAGTCKPDYVEINGEKYCGERSQFVYVS 420
DB 361 PPNIDCTWNIENVNNOHVKKRFKFFYLLEPGVAGTCKPDYVEINGEKYCGERSQFVYVS 420
QY 421 NSNKITVRFHSDQSYTDGTGFLAEYLSYDSDPCEGQFTCRGTGRCIRKELRCQGMADCTDH 480
DB 421 NSNKITVRFHSDQSYTDGTGFLAEYLSYDSDPCEGQFTCRGTGRCIRKELRCQGMADCTDH 480
QY 481 SDELINCSADAGHQTCKNKKFCCKPLFWVCDVNDGSDNSDEQSCSPAQTFFRCNSGKCLSK 540
DB 481 SDELINCSADAGHQTCKNKKFCCKPLFWVCDVNDGSDNSDEQSCSPAQTFFRCNSGKCLSK 540
QY 541 SQQNGKDDCGDSDSDASCPKNNVYTCRKYRCLNGICLSKNPEDCKEDCGSDGDEK 600
DB 541 SQQNGKDDCGDSDSDASCPKNNVYTCRKYRCLNGICLSKNPEDCKEDCGSDGDEK 600
QY 601 DCDGCLSFTRQAVVCGTDDEGEWPMQVSLHALGQGHICGASLISPNMLVSAHACYID 660
DB 601 DCDGCLSFTRQAVVCGTDDEGEWPMQVSLHALGQGHICGASLISPNMLVSAHACYID 660
QY 661 DRGRYSYDPTQWIAFLGLHDQSRSAPGVQRRILKRIISHPFNDFTFDYDIALLEBKP 720
DB 661 DRGRYSYDPTQWIAFLGLHDQSRSAPGVQRRILKRIISHPFNDFTFDYDIALLEBKP 720
QY 721 AEYSMWRTPICLPASHVFPKGAIMWTGMHTYGGGALILDKGEIRVINOTTENLL 780
DB 721 AEYSMWRTPICLPASHVFPKGAIMWTGMHTYGGGALILDKGEIRVINOTTENLL 780
QY 781 PQQITPRMVCVFLSGGVDSQCGSGGGLSVSEADGRIFGAGVSWGDCAGRNKPGVYT 840
DB 781 PQQITPRMVCVFLSGGVDSQCGSGGGLSVSEADGRIFGAGVSWGDCAGRNKPGVYT 840
QY 841 RLPLFRDMIKENTGV 855
DB 841 RLPLFRDMIKENTGV 855

```

```

RESULT 17
ABP56619
ID ABP56619 standard; protein, 855 AA.
XX
AC ABP56619;
XX
DT 24-MAR-2003 (first entry)
XX
DE Human membrane-type serine protease MTSP1 protein SEQ ID NO:2.
XX
KM Human: membrane-type serine protease; enzyme; MTSP10; cytostatic;
XX type-II membrane-type serine protease; neoplastic disease; tumour; MTSP1;
XX matrixase.
XX
OS Homo sapiens.
XX
PN MO200292841-A2.
XX
PD 21-NOV-2002.
XX
PF 14-MAY-2002; 2002WO-US015332.
XX
PR 14-MAY-2001; 2001US-0291001P.

```

```

XX
PA (CORV-) CORVAS INT INC.
XX
PI Madison EL, Yeh J;
XX
DR WPI; 2003-129309/12.
XX
DR N-PSDB; AB222450.
XX
PT New polypeptides comprising the protease domain of a type-II membrane-
PT type serine protease (MTCP10), or its mutants, useful for diagnosing
PT neoplasms or malignancies, or for screening for MTCP10 inhibitors for
PT treating such diseases.
XX
PS Disclosure; Page 181-183; 1988pp; English.
XX
CC The present invention describes a polypeptide comprising a purified
CC single or two chain polypeptide, which comprises the protease domain of a
CC type-II membrane-type serine protease (MTSP10) or its catalytically
CC active portion, or a mutant of it, where up to 50 % of the amino acids
CC are replaced with another amino acid, and the resulting polypeptide is a
CC single chain or two chain polypeptide that has a catalytic activity of at
CC least 1-10 % of the unmutated polypeptide. MTSP10 has cytostatic
CC activity. The polypeptide containing the protease domain of the MTSP10 is
CC useful for detecting a neoplastic disease, and for diagnosing the
CC presence of a pre-malignant lesion, a malignancy, or other pathologic
CC condition in a subject, or monitoring tumour (e.g. breast, cervix,
CC prostate, lung, ovary or colon tumour) progression and/or therapeutic
CC effectiveness. An inhibitor of the polypeptide containing the protease
CC domain of MTSP10 is useful for treating or preventing neoplastic disease
CC in a mammal. An inhibitor of the activation or cleavage of the zymogen form
CC of the MTSP10 polypeptide is useful for inhibiting tumour initiation.
CC growth or progression, or treating (pre-)malignant conditions of the e.g.
CC breast, cervix, prostate, lung, ovary or colon. The present sequence
CC represents human MTSP1 (also known as matrixase), which is used in an
CC example from the present invention
XX
SQ Sequence 855 AA:
Query Match 99.4%; Score 4653; DB 6; Length 855;
Best Local Similarity 99.5%; Pred. No. 2,4e-307;
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 MGSDBARKGGGKPRDPCAGLKYSRHKVNGLEBGEVFLPVNNVKKYKHPGRMVLAA 60
DB 1 MGSDBARKGGGKPRDPCAGLKYSRHKVNGLEBGEVFLPVNNVKKYKHPGRMVLAA 60
QY 61 VILGLLVILGIGFLVWHLQYRDVAVQKXNGYMRITNENFVDAYENSSTEBVSLASKY 120
DB 61 VILGLLVILGIGFLVWHLQYRDVAVQKXNGYMRITNENFVDAYENSSTEBVSLASKY 120
QY 121 KDALKILYSGVPLFGPYHKESAVTAPSEGSVIAIYNSSEFIPQHLVBEARVVAEERVVM 180
DB 121 KDALKILYSGVPLFGPYHKESAVTAPSEGSVIAIYNSSEFIPQHLVBEARVVAEERVVM 180
QY 181 LPPRABSLKSVTVSVVAFPTDSKTVOQTODNSCSFGIHAAGVELMRPTTGPDPSPVA 240
DB 181 LPPRABSLKSVTVSVVAFPTDSKTVOQTODNSCSFGIHAAGVELMRPTTGPDPSPVA 240
QY 241 HARCQMLRGDADSVLSLTFRSPDLASCDERGSDLVTVYNTLSPEMPEHALVOLCGTYPSP 300
DB 241 HARCQMLRGDADSVLSLTFRSPDLASCDERGSDLVTVYNTLSPEMPEHALVOLCGTYPSP 300
QY 301 YNLTFFSSQNVLLITLITNTERRHPGFEATFQOLPRMSSCGGLRKAQGTFFNSPYYPGHY 360
DB 301 YNLTFFSSQNVLLITLITNTERRHPGFEATFQOLPRMSSCGGLRKAQGTFFNSPYYPGHY 360
QY 361 PPNIDCTWNIENVNNOHVKKRFKFFYLLEPGVAGTCKPDYVEINGEKYCGERSQFVYVS 420
DB 361 PPNIDCTWNIENVNNOHVKKRFKFFYLLEPGVAGTCKPDYVEINGEKYCGERSQFVYVS 420
QY 421 NSNKITVRFHSDQSYTDGTGFLAEYLSYDSDPCEGQFTCRGTGRCIRKELRCQGMADCTDH 480
DB 421 NSNKITVRFHSDQSYTDGTGFLAEYLSYDSDPCEGQFTCRGTGRCIRKELRCQGMADCTDH 480

```

Qy	481	DELNCSDCADGHOFTCKNFKCPLFLVWCDSDVNDCCGNSDBOGGSCPAQOTFRCSNGKCLSK	540
Db	481	SDELNCSDCADGHOFTCKNFKCPLFLVWCDSDVNDCCGNSDBOGGSCPAQOTFRCSNGKCLSK	540
Qy	541	SQCGKXKDDCGGSDBEASCPRKNVVTCTGHTYRCCLNGCLCLSKSKNPNCDGKEKSDSDSK	600
Db	541	SQCGKXKDDCGGSDBEASCPRKNVVTCTGHTYRCCLNGCLCLSKSKNPNCDGKEKSDSDSK	600
Qy	601	DCDCGLRSFTQRARVVGCTDADDEGEMPMQVSLHALGQGHICGASLLISPMVLVSAAHCYID	660
Db	601	DCDCGLRSFTQRARVVGCTDADDEGEMPMQVSLHALGQGHICGASLLISPMVLVSAAHCYID	660
Qy	661	DRGFRTYDPTQMTATFLGLHDOSQSRSPVQOERLKRKIIISHPPENDTTFDYDIALLELEKX	720
Db	661	DRGFRTYDPTQMTATFLGLHDOSQSRSPVQOERLKRKIIISHPPENDTTFDYDIALLELEKX	720
Qy	721	AEYSMWBPICLPDASHVFPAGKAIWVTMGHTQYGGTALLIIQKEIRIYINOTTEMLL	780
Db	721	AEYSMWBPICLPDASHVFPAGKAIWVTMGHTQYGGTALLIIQKEIRIYINOTTEMLL	780
Qy	781	FOQITPRMWCVFLSGGVSDCOGDSGGLPSVEADRIFGAGVSWGDCAGRNKRGVYT	840
Db	781	FOQITPRMWCVFLSGGVSDCOGDSGGLPSVEADRIFGAGVSWGDCAGRNKRGVYT	840
Qy	841	RLPLFRDWIKENTGV	855
Db	841	RLPLFRDWIKENTGV	855

XX	RESULT 18
XX	AAO30146
XX	AAO30146 standard; protein; 855 AA.
XX	AAO30146;
XX	03-SEP-2003 (first entry)
XX	Human membrane-type serine protease MTSPI protein.
XX	Serine protease 17; CVPSP17; tumour; cancer; antisense therapy; prostate;
XX	breast; cervix; lung; ovary; colon; gene therapy; human; enzyme; MTSPI;
XX	membrane-type serine protease; matriptase.
XX	Homo sapiens.
XX	WO2003044179-A2.
XX	30-MAY-2003.
XX	20-NOV-2002; 2002MO-US037626.
XX	20-NOV-2001; 2001US-0332015P.
XX	(CORV-) CORVAS INT INC.
XX	Madison EL, Ong EO;
XX	WPI; 2003-449816/42.
XX	N-PSDB; AAL60792.
XX	New substantially purified serine protease 17 polypeptide and encoding
XX	nucleic acid; useful for diagnosing and treating tumor conditions and/or
XX	cancer, particularly of the breast, cervix, prostate, lung, ovary or
XX	colon.
XX	Disclosure; Page 175-177; 189pp; English.
XX	The invention relates to serine protease 17 polypeptide designated CVPSP17
XX	and its corresponding nucleic acid sequence. The invention also relates
XX	to a method using CVPSP17 protein to identify compounds that modulate its
XX	protease activity. The method is useful for preventing, diagnosing and
XX	treating disorders related to the serine protease 17 activity, such as

[illegible]

RESULT 19
AAE29820
ID AAE29820 standard; protein: 855 AA.
XX
AC AAE29820;
XX
DT 24-FEB-2003 (first entry)
XX
DE Human membrane-type serine protease 1 (MTSP1).
XX
KW Human; type II membrane-type serine protease 9; tumour; transgenic;
KW type II transmembrane serine protease; enzyme; gene therapy; MTSP9;
KW neoplastic disease; transgenic animal; membrane-type serine protease 1;
KW MTSP; MTSP1; matrilysin.
XX
OS Homo sapiens.
XX
PN WO200277267-A2.
XX
PD 03-OCT-2002.
XX
PF 27-MAR-2002; 2002WO-US009611.
XX
PR 27-MAR-2001; 2001US-0279228P.
PR 15-MAY-2001; 2001US-0291501P.
XX
PA (CORV-) CORVAS INT INC.
PI Madison EL, Ong EO;
XX
DR WPI: 2003-018940/01.
DR N-PSDB; AAD47225.
XX
PT New substantially purified single or two-chain type II membrane-type
PT serine protease 9 (MTSP9) polypeptide, useful for monitoring tumor
PT progression, inhibiting tumor initiation, or treating a malignant or pre-
PT malignant condition.
XX
PS Disclosure; Page 183-185; 199pp; English.
XX
CC The invention relates to type II membrane-type serine protease 9 (MTSP9)
CC polypeptides and polynucleotides. MTSP belongs to type II transmembrane
CC serine protease (TSP) family. Sequences of the invention and their
CC antibodies are useful for diagnosing, treating or preventing neoplastic
CC disease in mammals. They are useful for monitoring tumor progression,
CC inhibiting tumor initiation, growth or progression or treating malignant
CC or pre-malignant conditions. Transgenic animals of the invention are
CC useful in animal models of tumor initiation, growth and/or progression
CC models. The invention is also useful in gene therapy. The present
CC sequence is human membrane-type serine protease 1 (MTSP1). MTSP1 also
CC referred as matrilysin is a member of the TSP family
XX
SQ Sequence 855 AA;

Query Match 99.4%; Score 4653; DB 6; Length 855;
Best Local Similarity 99.5%; Pred. No. 2.4e-307;
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 MGSDRAKGGGGPGDFAGLKYNSRHEKYNGLAEVEFLPVNNYKVEKHGPGHVVLA 60
DB 1 MGSDRAKGGGGPGDFAGLKYNSRHEKYNGLAEVEFLPVNNYKVEKHGPGHVVLA 60
QY 61 VLIGLLLVLLIGLGLVWHLQYRDYRVQKVKNGYKRNITNENFVDAVENSNGTEFVSLASKV 120
DB 61 VLIGLLLVLLIGLGLVWHLQYRDYRVQKVKNGYKRNITNENFVDAVENSNGTEFVSLASKV 120
QY 121 KDALKLTVSGVPLGIPYHKEASVTAFAFSGSVIAYWMSFSLPOHLVEBAEAVMBEERVVM 180
DB 121 KDALKLTVSGVPLGIPYHKEASVTAFAFSGSVIAYWMSFSLPOHLVEBAEAVMBEERVVM 180
QY 121 KDALKLTVSGVPLGIPYHKEASVTAFAFSGSVIAYWMSFSLPOHLVEBAEAVMBEERVVM 180
DB 121 KDALKLTVSGVPLGIPYHKEASVTAFAFSGSVIAYWMSFSLPOHLVEBAEAVMBEERVVM 180
QY 181 LPPARSLKSFVTVSVVAFPTDSKTVCRTQDNCSCFGLHARGVELMRTTGPFPDSPYPA 240
DB 181 LPPARSLKSFVTVSVVAFPTDSKTVCRTQDNCSCFGLHARGVELMRTTGPFPDSPYPA 240
QY 181 LPPARSLKSFVTVSVVAFPTDSKTVCRTQDNCSCFGLHARGVELMRTTGPFPDSPYPA 240
DB 181 LPPARSLKSFVTVSVVAFPTDSKTVCRTQDNCSCFGLHARGVELMRTTGPFPDSPYPA 240

QY 241 HARCQMALRGDADSVLSTFRSFDLASCDERGSDLVTYVNTLSPMEPHALVOLCGTYPPS 300
DB 241 HARCQMALRGDADSVLSTFRSFDLASCDERGSDLVTYVNTLSPMEPHALVOLCGTYPPS 300
QY 301 YNLTFFSSQNVLLITLITNTERRHPGEATFPLPRMSCGGRLLRKAQGTFSPPYBGHY 360
DB 301 YNLTFFSSQNVLLITLITNTERRHPGEATFPLPRMSCGGRLLRKAQGTFSPPYBGHY 360
QY 361 PENIDCTWNIIEVPPNQHVKRFRKFFYLLEPGVAGTCPKDYVBIINGEKYCGERSQFVYTS 420
DB 361 PENIDCTWNIIEVPPNQHVKRFRKFFYLLEPGVAGTCPKDYVBIINGEKYCGERSQFVYTS 420
QY 421 NSNKITVRFHSDDSYDTGTFLAEVLSYSSDDPCGQFTCTGRCIRKELACDGMADCTDH 480
DB 421 NSNKITVRFHSDDSYDTGTFLAEVLSYSSDDPCGQFTCTGRCIRKELACDGMADCTDH 480
QY 481 SDELMSCDAGHOFCTCKNFKCKPLFWVCDSVNDGSDNDEQSCSPAQTFRCNSGKCLSK 540
DB 481 SDELMSCDAGHOFCTCKNFKCKPLFWVCDSVNDGSDNDEQSCSPAQTFRCNSGKCLSK 540
QY 541 SQQCNKGKDCGDSDEASCPKRVNVVCTTKTTRCLNGLCLSKNPNCEGKEDSDGSEK 600
DB 541 SQQCNKGKDCGDSDEASCPKRVNVVCTTKTTRCLNGLCLSKNPNCEGKEDSDGSEK 600
QY 601 DCCGGLRSPTRQARVVGCTADDEGEMPMOVSLLHALGGGHI CGASLLSPMVLVSAHCTID 660
DB 601 DCCGGLRSPTRQARVVGCTADDEGEMPMOVSLLHALGGGHI CGASLLSPMVLVSAHCTID 660
QY 661 DRGFRYSPTQMTAFGLIHDQSOSAAGVQERRLKRIISHPFNDFTFDYDIALLELEKP 720
DB 661 DRGFRYSPTQMTAFGLIHDQSOSAAGVQERRLKRIISHPFNDFTFDYDIALLELEKP 720
QY 721 AEYSWVRPICLPDASHVFPAGKAIWVGHTQYGGTALLIOKGIIRVYNQTCENL 780
DB 721 AEYSWVRPICLPDASHVFPAGKAIWVGHTQYGGTALLIOKGIIRVYNQTCENL 780
QY 781 PQQITPRMVCVFLSGVSDSCGDSGGPLSSVADGRIFGAGVSWMGCGAGRNKPGVYT 840
DB 781 PQQITPRMVCVFLSGVSDSCGDSGGPLSSVADGRIFGAGVSWMGCGAGRNKPGVYT 840
QY 841 RLPLFRDMIKENTGV 855
DB 841 RLPLFRDMIKENTGV 855

RESULT 20
AAE29791
ID AAE29791 standard; protein: 855 AA.
XX
AC AAE29791;
XX
DT 24-FEB-2003 (first entry)
XX
DE Human membrane-type serine protease, MTSP1.
XX
KW Human; serine protease 14; CYP14; cancer; malignancy; breast; colon;
KW gene therapy; cytostatic; membrane-type serine protease; enzyme; MTSP1.
XX
OS Homo sapiens.
XX
PN WO200277263-A2.
XX
PD 03-OCT-2002.
XX
PF 20-MAR-2002; 2002WO-US009039.
XX
PR 22-MAR-2001; 2001US-0278166P.
XX
PA (CORV-) CORVAS INT INC.
PI Madison EL, Yeh J;
XX

DR WPI, 2003-018938/01.
 DR N-PSDB; AAD47180.
 XX New purified CVSP14 polypeptide and encoding nucleic acid molecule.
 PT useful for diagnosing, preventing and/or treating disorders, such as
 PT cancers and malignancies of the breast, cervix, prostate, lung, ovary or
 PT colon.
 XX
 XX Disclosure; Page 171-173; 185pp; English.
 PS
 CC The invention relates to transmembrane serine protease 14 (CVSP14), its
 CC nucleic acid sequence and the method based on them. The methods and
 CC compositions of the invention are useful for diagnosing, preventing
 CC and/or treating conditions associated with the aberrant expression or
 CC activity of the CVSP14 polypeptide, such as cancers and malignancies of
 CC the breast, cervix, prostate, lung, ovary or colon. The methods are also
 CC useful for identifying compounds that will modulate the protease activity
 CC of CVSP14 polypeptide, and monitoring tumour progression and/or
 CC therapeutic effectiveness. CVSP14 DNA used in gene therapy. The present
 CC sequence is human membrane-type serine protease, MTSP1
 CC
 XX
 XX Sequence 855 AA:
 SQ
 Query Match 99.4%; Score 4653; DB 6; Length 855;
 Best Local Similarity 99.5%; Pred. No. 2,4e-307;
 Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 MGSDRAKGGGKPGAGLKYNSRHKVNGLEGEVEFLPVNNVKYKXGPGRWVLLAA 60
 DB 1 MGSDRAKGGGKPGAGLKYNSRHKVNGLEGEVEFLPVNNVKYKXGPGRWVLLAA 60
 QY 61 VLIGLLVLLIGLGVVHLYQYRDVAVQKXNGYMRITNENFVADYENSNSTEFLVSLASKV 120
 DB 61 VLIGLLVLLIGLGVVHLYQYRDVAVQKXNGYMRITNENFVADYENSNSTEFLVSLASKV 120
 QY 121 KDALKLLYSGVPTGPRHKSAAVTAPESEGVIAAYNSEFSLPQHLVEAEKVAEERVVM 180
 DB 121 KDALKLLYSGVPTGPRHKSAAVTAPESEGVIAAYNSEFSLPQHLVEAEKVAEERVVM 180
 QY 181 LPPPARSLKSFVVTSSVAAPFTDSKTVOPTODNSCSFGLHARGVLMFTTGPFDSPYPA 240
 DB 181 LPPPARSLKSFVVTSSVAAPFTDSKTVOPTODNSCSFGLHARGVLMFTTGPFDSPYPA 240
 QY 241 HARCQMLRGDADSVLSLTFRSPDLASCDERGSOLVTYNTLSPEPHALVOLCGTYPSP 300
 DB 241 HARCQMLRGDADSVLSLTFRSPDLASCDERGSOLVTYNTLSPEPHALVOLCGTYPSP 300
 QY 301 YNLTFHSSQNVLLITLTNTERRHPRGEATFPQLPRMSSCGRLRKAQGTFSNPPYFGHY 360
 DB 301 YNLTFHSSQNVLLITLTNTERRHPRGEATFPQLPRMSSCGRLRKAQGTFSNPPYFGHY 360
 QY 361 PPNDICWNIIEVPNNQHVAFKFFVYLEPGVPAQTCKDVEINGEKYGCSERQFVVTSS 420
 DB 361 PPNDICWNIIEVPNNQHVAFKFFVYLEPGVPAQTCKDVEINGEKYGCSERQFVVTSS 420
 QY 421 NSNKITVRFHSDQSYDTDTGLAEVLSYDSDPCGQFTCRGRCIRKELRCDGNADCTDH 480
 DB 421 NSNKITVRFHSDQSYDTDTGLAEVLSYDSDPCGQFTCRGRCIRKELRCDGNADCTDH 480
 QY 481 SDEINSCDAGHOTCTCKNRCCKPLFWICDSVNDGDSDEGSCCPAQTFRCSNGKCLSK 540
 DB 481 SDEINSCDAGHOTCTCKNRCCKPLFWICDSVNDGDSDEGSCCPAQTFRCSNGKCLSK 540
 QY 541 SQQNGNGDGDGSDGDEASCPKVNVTCTKHYRCLANGCLSKNGPECDGKEDCSGSDGDEK 600
 DB 541 SQQNGNGDGDGSDGDEASCPKVNVTCTKHYRCLANGCLSKNGPECDGKEDCSGSDGDEK 600
 QY 601 DCDGGLRSFTRQARVVGCTADABEGEMWQVSLHALGQCHTIGASLISPMNLVSAHACYID 660
 DB 601 DCDGGLRSFTRQARVVGCTADABEGEMWQVSLHALGQCHTIGASLISPMNLVSAHACYID 660
 QY 661 DRGFRYSDPTQWTAFLGLHDQSQRSAFGVQERLRKRIISHPFNDPTFEDVIALLELEKP 720
 DB 661 DRGFRYSDPTQWTAFLGLHDQSQRSAFGVQERLRKRIISHPFNDPTFEDVIALLELEKP 720

DB 661 DRGFRYSDPTQWTAFLGLHDQSQRSAFGVQERLRKRIISHPFNDPTFEDVIALLELEKP 720
 QY 721 AEYSSWVRPCLDPADSHVFPAGKAIWVTGHTQYGGTGLILQKGEIRVYNQTCENLL 780
 DB 721 AEYSSWVRPCLDPADSHVFPAGKAIWVTGHTQYGGTGLILQKGEIRVYNQTCENLL 780
 QY 781 PQQITPRMVCVGLSGVDSCQDGGGGLSSVADGRIFGAGVVSMDGCAGNKRGVYT 840
 DB 781 PQQITPRMVCVGLSGVDSCQDGGGGLSSVADGRIFGAGVVSMDGCAGNKRGVYT 840
 QY 841 RLPLFRDWIKENTGV 855
 DB 841 RLPLFRDWIKENTGV 855
 RESULT 21
 ID ABP72376 standard; protein; 855 AA.
 AC ABP72376;
 DT 13-MAY-2003 (first entry)
 DE Transmembrane serine protease 1 (MTSP1).
 XX Human; transmembrane serine protease 1; MTSP1; MTSP20; enzyme;
 KM cytosolic; dermatological; cardiac; vulnary; ophthalmological;
 KM gene therapy.
 XX Homo sapiens.
 OS
 PN WC020304681-A2.
 XX
 XX 16-JAN-2003.
 PR 03-JUL-2002; 2002MO-US021208.
 XX
 PR 03-JUL-2001; 2001US-0302939P.
 XX
 PA (CORV-) CORVAS INT INC.
 PI Madison EL, Ong EO;
 XX
 XX WPI, 2003-239207/23.
 DR N-PSDB; ABZ58500.
 XX
 PT New type-II membrane-type serine protease 20 polypeptides, useful for
 PT preparing a medical treatment for diagnosing, treating or preventing cancer,
 PT dermatological disorders, aberrant wound repairs or crest syndromes.
 PS
 XX Disclosure; Page 198-200; 216pp; English.
 XX
 XX The present sequence is the protein sequence of human type II
 XX transmembrane serine protease 1 (MTSP1). The invention relates to novel
 XX human type II transmembrane serine protease 20 (MTSP20) (see ABP72374) and
 XX nucleic acids encoding it (see ABZ58499). Also claimed are methods of
 XX inhibiting tumour initiation, growth or progression by inhibiting MTSP20
 XX activity, and of treating or preventing a disease or disorder associated
 XX with undesired and/or uncontrolled angiogenesis or neovascularisation,
 XX especially undesired angiogenesis associated with solid neoplasms,
 XX vascular malformations and cardiovascular disorders, chronic inflammatory
 XX diseases, aberrant wound repairs, circulatory disorders, crest syndromes,
 XX dermatological disorders and ocular disorders using an inhibitor of MTSP20
 XX
 XX Sequence 855 AA:
 SQ
 Query Match 99.4%; Score 4653; DB 6; Length 855;
 Best Local Similarity 99.5%; Pred. No. 2,4e-307;
 Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 MGSDRAKGGGKPGAGLKYNSRHKVNGLEGEVEFLPVNNVKYKXGPGRWVLLAA 60
 DB 1 MGSDRAKGGGKPGAGLKYNSRHKVNGLEGEVEFLPVNNVKYKXGPGRWVLLAA 60

QY VLIGLLVLIGIGLVVHLYRDVYVOKVKNYRITNENFVDAYENSNSTEFLVLSKV 120
DB VLIGLLVLIGIGLVVHLYRDVYVOKVKNYRITNENFVDAYENSNSTEFLVLSKV 120
QY KDALKLISYGVPLGPIYHKESAVTAFSEGSVIAYWSEFSIPQHLVEAEARVMAEERVVM 180
DB KDALKLISYGVPLGPIYHKESAVTAFSEGSVIAYWSEFSIPQHLVEAEARVMAEERVVM 180
QY LPPARSLKSFVVTSVVAFPTDSKTVOQTQDNCSFGLHARGVLMFTTGPFDSPYPA 240
DB LPPARSLKSFVVTSVVAFPTDSKTVOQTQDNCSFGLHARGVLMFTTGPFDSPYPA 240
QY HARCOMALRGDADSVLSLTFRSFDLASCDESGSLVTYNTLSMPEHALVOLGCTYPPS 300
DB HARCOMALRGDADSVLSLTFRSFDLASCDESGSLVTYNTLSMPEHALVOLGCTYPPS 300
QY YNLTFHSSQNVLLITLTNTERRHPGFEATFPQLPRMSSCGRLRKAQGTFSNPPYPGHY 360
DB YNLTFHSSQNVLLITLTNTERRHPGFEATFPQLPRMSSCGRLRKAQGTFSNPPYPGHY 360
QY PENIDCTWNIEVPNNQHVKVSFKFFYLLEPGVPAQTCPKDYVEINGEKCYCGERSQFVVT 420
DB PENIDCTWNIEVPNNQHVKVSFKFFYLLEPGVPAQTCPKDYVEINGEKCYCGERSQFVVT 420
QY NSNKITVRPHSDOSYDTDTGFLAEYLSYSDPCPGQFTCRTRGCIARKELRCDGMADCTDH 480
DB NSNKITVRPHSDOSYDTDTGFLAEYLSYSDPCPGQFTCRTRGCIARKELRCDGMADCTDH 480
QY SDELINCSGDAHQFTCKNKKFCCKPLFWVCDVNDGSDNSDEGSCCPAQTFRCNSGKCLSK 540
DB SDELINCSGDAHQFTCKNKKFCCKPLFWVCDVNDGSDNSDEGSCCPAQTFRCNSGKCLSK 540
QY SQQNGKXDDCGSDGDEASCPKRVNVTCTKHTYRCLNGICLSKGNPECDGKEDCSGSDSK 600
DB SQQNGKXDDCGSDGDEASCPKRVNVTCTKHTYRCLNGICLSKGNPECDGKEDCSGSDSK 600
QY DCDGGLSFTROAVVVGSTDADEBEMPVQSLHLAGQHGICGASLISPMVLVSAHACYID 660
DB DCDGGLSFTROAVVVGSTDADEBEMPVQSLHLAGQHGICGASLISPMVLVSAHACYID 660
QY DRGRYSDPTQMTAFGLIHDSORSAPGQBRRLKRIISHPPFNDFTFYDIALLELEKP 720
DB DRGRYSDPTQMTAFGLIHDSORSAPGQBRRLKRIISHPPFNDFTFYDIALLELEKP 720
QY AEYSMWVRPCLPASHVFPAGKAIWTTGHTOYGGTGAIILOKGEIRVINOTTCEML 780
DB AEYSMWVRPCLPASHVFPAGKAIWTTGHTOYGGTGAIILOKGEIRVINOTTCEML 780
QY POQITPRMVCVGLSGVDSGCGSGGLSSVEADGRIFGAGVWSWGGCGAGRMKPGVYT 840
DB POQITPRMVCVGLSGVDSGCGSGGLSSVEADGRIFGAGVWSWGGCGAGRMKPGVYT 840
QY RLPLFRDMIKENTGV 855
DB RLPLFRDMIKENTGV 855
RESULT 22
ADB97551 ID ADB97551 standard; protein; 855 AA.
XX ADB97551;
XX 04-DEC-2003 (first entry)
DE Human MTSpl, SEQ ID NO:2.
XX Human; type I transmembrane serine protease 1; MTSpl; matrix; serine
XX beeline protease; protease cleavage activation; diagnostic marker;
XX neoplastic disorder; cancer; breast; cervix; prostate; lung; ovary;
XX cytosolic; gene therapy; drug screening; tumour progression; monitoring;
XX enzyme.

XX Homo sapiens.
OS
XX WO2003031585-A2.
PN
XX 17-APR-2003.
PD
XX
PF 08-OCT-2002; 2002MO-US032417.
XX
PR 09-OCT-2001; 2001US-0328530P.
XX
PA (CORV-) CORVAS INT INC.
XX
PI Madison EL, Yeh J;
XX WPI; 2003-393442/37.
DR N-PSDB; ADB97550.
XX
XX New purified single- or two-chain polypeptide, useful for diagnosing,
PT preventing or treating cancer (e.g. colon cancer), comprises a protease
PT domain of a type-1 membrane-type serine protease 25 or its catalytically
PT active portion.
XX
XX Disclosure; Page 179-181; 97pp; English.
XX
XX The invention relates to human type I transmembrane serine protease 25
CC (MTSP25; ADB97555) and polypeptides derived from it (e.g., ADB97555). The
CC MTSP25 gene is located on chromosome 12. MTSP25 is a serine protease
CC which is activated by protease cleavage to yield a two-chain protease
CC comprising an A chain and a B chain linked by a disulphide bond. MTSP25
CC is expressed or is active in tumour cells, and can therefore be used as a
CC diagnostic marker for certain cancers. The invention also encompasses
CC nucleic acids encoding an MTSP25 polypeptide (ADB97564, ADB97571,
CC ADB97554); nucleic acid vectors and host cells comprising an MTSP25
CC polynucleotide; a MTSP25 knockout animal; and an antibody specific for
CC either the single chain (zymogen) or two-chain (activated) form of
CC MTSP25. MTSP25 polypeptides are useful in diagnosing, preventing or
CC treating neoplastic diseases, such as cancer of the breast, cervix,
CC prostate, lung, ovary or colon. MTSP25 polypeptides may also be used for
CC identifying compounds that modulate the protease activity of the
CC polypeptide and for monitoring tumour progression and/or therapeutic
CC effectiveness. The present sequence represents the related protein, MTSpl
CC (also known as matrilysin).
XX
SQ Sequence 855 AA;
Query Match 99.4%; Score 4653; DB 7; Length 855;
Best Local Similarity 99.5%; Pred. No. 2,4e-307;
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 MGSDRARKGGGKDPFGAGLKYSRHEKNGLEGEVFLPVNNYKYEKGPRWVLLAA 60
DB 1 MGSDRARKGGGKDPFGAGLKYSRHEKNGLEGEVFLPVNNYKYEKGPRWVLLAA 60
QY VLIGLLVLIGIGLVVHLYRDVYVOKVKNYRITNENFVDAYENSNSTEFLVLSKV 120
DB VLIGLLVLIGIGLVVHLYRDVYVOKVKNYRITNENFVDAYENSNSTEFLVLSKV 120
QY KDALKLISYGVPLGPIYHKESAVTAFSEGSVIAYWSEFSIPQHLVEAEARVMAEERVVM 180
DB KDALKLISYGVPLGPIYHKESAVTAFSEGSVIAYWSEFSIPQHLVEAEARVMAEERVVM 180
QY LPPARSLKSFVVTSVVAFPTDSKTVOQTQDNCSFGLHARGVLMFTTGPFDSPYPA 240
DB LPPARSLKSFVVTSVVAFPTDSKTVOQTQDNCSFGLHARGVLMFTTGPFDSPYPA 240
QY HARCOMALRGDADSVLSLTFRSFDLASCDESGSLVTYNTLSMPEHALVOLGCTYPPS 300
DB HARCOMALRGDADSVLSLTFRSFDLASCDESGSLVTYNTLSMPEHALVOLGCTYPPS 300
QY YNLTFHSSQNVLLITLTNTERRHPGFEATFPQLPRMSSCGRLRKAQGTFSNPPYPGHY 360
DB YNLTFHSSQNVLLITLTNTERRHPGFEATFPQLPRMSSCGRLRKAQGTFSNPPYPGHY 360

QY 361 PPNIIDCTWNIIEVNNQHVKVRFEFFYLLEPEVPAGTCPKDYVEINGEKYCGERSQFVVT
DB 361 PPNIIDCTWNIIEVNNQHVKVSFKFFYLLEPEVPAGTCPKDYVEINGEKYCGERSQFVVT
QY 421 NSNKITVRFHSDSYTDTGFLAEYLSYDSDPGQFTCTGRCIRKELRCDGMADCTDH
DB 421 NSNKITVRFHSDSYTDTGFLAEYLSYDSDPGQFTCTGRCIRKELRCDGMADCTDH
QY 481 SDELNCSDAAGHQTCKNKKCKPLFWVCDVNDGDNDSDEQSCSPAQTFRCSNGKCLSK
DB 481 SDELNCSDAAGHQTCKNKKCKPLFWVCDVNDGDNDSDEQSCSPAQTFRCSNGKCLSK
QY 541 SQQNGKDDCGDSDASCPKVVVVTCTKTYRCLNGLCLSKNPECDGKEDSDSDEK
DB 541 SQQNGKDDCGDSDASCPKVVVVTCTKTYRCLNGLCLSKNPECDGKEDSDSDEK
QY 601 DCCGGLRSFTROARVVGTTADDEGEMPMQVSLHALGQGHICGASLISPMVLVSAHCYID
DB 601 DCCGGLRSFTROARVVGTTADDEGEMPMQVSLHALGQGHICGASLISPMVLVSAHCYID
QY 661 DRGFRYSPTQMTAFGLGHDQSORAPGVQBERLKRITISHPFNDFTFDIALLELKP
DB 661 DRGFRYSPTQMTAFGLGHDQSORAPGVQBERLKRITISHPFNDFTFDIALLELKP
QY 721 AEVSSWVRPCLPDASHVFPAGKAIWVTGMGHTQYGGTGALLIQKGIRVINQTTCEML
DB 721 AEVSSWVRPCLPDASHVFPAGKAIWVTGMGHTQYGGTGALLIQKGIRVINQTTCEML
QY 781 PQQITPRMNCVGLSGVDSQCGSGPLSSVEADGRIFGAGVSVWGDGCAGRKPEVYT
DB 781 PQQITPRMNCVGLSGVDSQCGSGPLSSVEADGRIFGAGVSVWGDGCAGRKPEVYT
QY 841 RLPLFRDMIKENTGV 855
DB 841 RLPLFRDMIKENTGV 855

RESULT 23
AD10371
ID AD10371 standard; protein; 855 AA.
XX
AC AD10371;
XX
DT 22-APR-2004 (first entry)
XX
DE Human cell surface protease #1.
XX
KW therapeutic agent; plasmin; protease specific antigen; PSA;
KW cell-surface protease-associated disease; cancer; ocular disease;
KW cardiovascular disease; chronic inflammatory disease; wound;
KW circulatory disorder; dermatological disorder; rheumatoid arthritis;
KW psoriasis; diabetic retinopathy; pterygium;
KW excimer laser surgery scarring; glaucoma filtering surgery scarring;
KW macular degeneration; cleft syndrome; solid neoplasia; vascular tumour;
KW melanoma; Kaposi's sarcoma; human; cell surface protease.
XX
OS Homo sapiens.
XX
PN WO200295007-A2.
XX
PD 28-NOV-2002.
XX
PF 23-MAY-2002; 2002MO-US016819.
XX
PR 23-MAY-2001; 2001US-0293267P.
XX
PA (CORV-) CORVAS INT INC.
XX
PI Madison EL, Semple JE, Vlaauk GP, Kemp SJ, Komandla M, Stev DV;
XX WPI; 2003-221280/21.
DR N-PSDB; AD10370.

XX
PT Novel conjugate useful for treating cell-surface protease-associated
PT disease, comprises a therapeutic agent and a peptidic or nucleic acid
PT substrate linked to it optionally by a peptidic linker.
XX
PS Claim 9; SEQ ID NO 2; 581bp; English.
XX
CC The invention comprises a conjugate that consists of a therapeutic agent
CC and a peptide substrate (optionally linked via linker). The peptide
CC substrate is proteolytically cleaved by a cell surface protease or a
CC soluble, released or shed form of it, to liberate the therapeutic agent,
CC the conjugate of the invention is not substantially cleaved by plasmin or
CC protease specific antigen (PSA). The conjugate of the invention is useful
CC for treating a cell-surface protease-associated disease such as: cancer,
CC ocular diseases, cardiovascular diseases, chronic inflammatory diseases,
CC wound, circulatory disorders, dermatological disorders, rheumatoid
CC arthritis, psoriasis, diabetic retinopathy, recurrence of pterygium,
CC scarring from excimer laser surgery, scarring from glaucoma filtering
CC surgery, macular degeneration, cleft syndrome, solid neoplasia, vascular
CC tumours, melanoma and Kaposi's sarcoma. The present amino acid sequence
CC represents a human cell surface protease.
XX
SQ Sequence 855 AA.
Query Match 99.4%; Score 4653; DB 7; Length 855;
Best Local Similarity 99.5%; Pred. No. 2, 4e-307;
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 MSDDRARKGGGPKDAGAGIKVNSRHEKVGLEEGVEFLPVNNVKKYKPKRWVTLAA 60
DB 1 MSDDRARKGGGPKDAGAGIKVNSRHEKVGLEEGVEFLPVNNVKKYKPKRWVTLAA 60
QY 61 VILGLLVILGIFLVNHLQYRDVRYQKVNGYMRITNENFVDAYENSTEFVSLASKV 120
DB 61 VILGLLVILGIFLVNHLQYRDVRYQKVNGYMRITNENFVDAYENSTEFVSLASKV 120
QY 121 KDALKLLYSGVPLRGPHKESAVTAPSEGSVIAIYNSERSIPQHLVBEARVMAERVVM 180
DB 121 KDALKLLYSGVPLRGPHKESAVTAPSEGSVIAIYNSERSIPQHLVBEARVMAERVVM 180
QY 181 LPPRARSLSFVVTSVVAPPTDSKTQVORODNSCSGLHARGVLEMRFTTGGPPDSPYA 240
DB 181 LPPRARSLSFVVTSVVAPPTDSKTQVORODNSCSGLHARGVLEMRFTTGGPPDSPYA 240
QY 241 HARCQALRGDADSVSLTFRSFDLASCDERGSIDLTVYNTLSLPMEPHALVOLCGTYP 300
DB 241 HARCQALRGDADSVSLTFRSFDLASCDERGSIDLTVYNTLSLPMEPHALVOLCGTYP 300
QY 301 YNLTFFSSQNVLLITLITNTERHPGEATFPOLPMNSCGGRUKRAQGTFSPTPGHY 360
DB 301 YNLTFFSSQNVLLITLITNTERHPGEATFPOLPMNSCGGRUKRAQGTFSPTPGHY 360
QY 361 PPNIIDCTWNIIEVNNQHVKVRFEFFYLLEPEVPAGTCPKDYVEINGEKYCGERSQFVVT 420
DB 361 PPNIIDCTWNIIEVNNQHVKVSFKFFYLLEPEVPAGTCPKDYVEINGEKYCGERSQFVVT 420
QY 421 NSNKITVRFHSDSYTDTGFLAEYLSYDSDPGQFTCTGRCIRKELRCDGMADCTDH 480
DB 421 NSNKITVRFHSDSYTDTGFLAEYLSYDSDPGQFTCTGRCIRKELRCDGMADCTDH 480
QY 481 SDELNCSDAAGHQTCKNKKCKPLFWVCDVNDGDNDSDEQSCSPAQTFRCSNGKCLSK 540
DB 481 SDELNCSDAAGHQTCKNKKCKPLFWVCDVNDGDNDSDEQSCSPAQTFRCSNGKCLSK 540
QY 541 SQQNGKDDCGDSDASCPKVVVVTCTKTYRCLNGLCLSKNPECDGKEDSDSDEK 600
DB 541 SQQNGKDDCGDSDASCPKVVVVTCTKTYRCLNGLCLSKNPECDGKEDSDSDEK 600
QY 601 DCCGGLRSFTROARVVGTTADDEGEMPMQVSLHALGQGHICGASLISPMVLVSAHCYID 660
DB 601 DCCGGLRSFTROARVVGTTADDEGEMPMQVSLHALGQGHICGASLISPMVLVSAHCYID 660
QY 661 DRGFRYSPTQMTAFGLGHDQSORAPGVQBERLKRITISHPFNDFTFDIALLELKP 720

Db 661 DRGRYSDPTQWTFALGLHDQSASAPGVQERRLKRIISHPFDFTFDDIALLELEXP 720
Qy 721 AEYSWVRPCLPDASHVFPAGKAIWTTGHTGYGTGALLQKGEIRVINQTTCENL 780
Db 721 AEYSWVRPCLPDASHVFPAGKAIWTTGHTGYGTGALLQKGEIRVINQTTCENL 780
Qy 781 PQQITPRMCCVGLSGVDSQCGSGPLSSVEADGRIFGAGVVSMDGCGAKRKPQVYT 840
Db 781 PQQITPRMCCVGLSGVDSQCGSGPLSSVEADGRIFGAGVVSMDGCGAKRKPQVYT 840
Qy 841 RLPLFRDMIKENTGV 855
Db 841 RLPLFRDMIKENTGV 855
RESULT 24
ADG65326
ID ADG65326 standard; protein; 855 AA.
AC ADG65326;
XX 11-MAR-2004 (first entry)
XX DE Human MTSPI.
XX human, type II membrane-type serine protease 12; MTSPI2; chromosome 19;
KM protease domain; MTSPI2-P1; MTSPI2-P2; MTSPI2-P3; lung; oesophageal;
KM prostate; colon; ovary; cervix; breast; pancreas; cancer; tumour;
XX neoplastic condition.
XX Homo sapiens.
XX Key Location/Qualifiers
FH 615..855
FT Domain /label = Protease domain
XX W02003104394-A2.
XX 18-DEC-2003.
XX 21-MAY-2003; 2003WO-US016181.
XX 21-MAY-2002; 2002US-0382851P.
XX (DEND-) DENDREON SAN DIEGO LLC.
XX Madison EL, Ong EO;
XX WPI; 2004-062325/06.
XX New type II membrane-type serine protease 12 proteins and nucleic acids,
PT useful in diagnostics, particularly for diagnosing lung, oesophageal,
XX prostate, colon, ovary, cervix, breast and pancreas cancers.
XX Disclosure; SEQ ID NO 2; 207bp; English.
XX This sequence represents a human type II membrane-type serine protease 1
CC (MTSPI). The MTSPI protein sequence was used in the isolation of the
CC coding sequence of MTSPI2, which has been isolated to chromosome 19.
CC MTSPI2 does not include the sequence of amino acids Arg-1ys-His-Ieu-Pro-
CC Arg-Pro-Ala ADG65347. The MTSPI2 coding sequence was identified by using
CC the protein sequence of the protease domain of MTSPI to search the human
CC HTGS database which produced three serine proteases, MTSPI2-P1, MTSPI2-
CC P2 and MTSPI2-P3. EST's were identified which matched portions of
CC MTSPI2-P1, MTSPI2-P2 and MTSPI2-P3. MTSPI2 polypeptides, proteins and
CC nucleic acids are useful in diagnostics, particularly for diagnosing
CC lung, oesophageal, prostate, colon, ovary, cervix, breast and pancreas
CC cancers. These are useful in immunoassays to detect, prognose, diagnose,
CC or monitor various conditions, diseases, and disorders affecting MTSPI2
CC polypeptide expression, or monitor their treatment. Modulators of MTSPI2
CC are useful for treating cancer, tumour and other neoplastic conditions.

Seq Sequence 855 AA;
Query Match 99.4%; Score 4653; DB 8; Length 855;
Best Local Similarity 99.5%; Pred. No. 2,4e-307;
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 1 MGSDBARKGGGKPDGAGLKYNSRHEKVNGLSEGVFLPVNNVKYKKEKPGKRWVLA 60
Db 1 MGSDBARKGGGKPDGAGLKYNSRHEKVNGLSEGVFLPVNNVKYKKEKPGKRWVLA 60
Qy 61 VLIGLLVLLIGLIGLVHMLQYRDVRVQKXNGYRITNENFVDAYENSSTEFYSLASKV 120
Db 61 VLIGLLVLLIGLIGLVHMLQYRDVRVQKXNGYRITNENFVDAYENSSTEFYSLASKV 120
Qy 121 KDALKLYSGVPLGPRHYKESAATAPSEGSVIAYYWSEFSIPQHLVEARVMAEERVM 180
Db 121 KDALKLYSGVPLGPRHYKESAATAPSEGSVIAYYWSEFSIPQHLVEARVMAEERVM 180
Qy 181 LPPRARSLSKFVVTSSVVAAPPTDSKTVQRTODNSCSFGLHARGVELMFTTGPFPDSPYA 240
Db 181 LPPRARSLSKFVVTSSVVAAPPTDSKTVQRTODNSCSFGLHARGVELMFTTGPFPDSPYA 240
Qy 241 HARCQVALRGDADSVLSLTFRSFDLASCDEGSDLVTVNTLSPEMHALVOLCGTYPPS 300
Db 241 HARCQVALRGDADSVLSLTFRSFDLASCDEGSDLVTVNTLSPEMHALVOLCGTYPPS 300
Qy 301 YNLTFFSSQVNLITLITNTERRHPRGEATFPQLPRMSGCCGRRLRKAQCTFNSPYRGHY 360
Db 301 YNLTFFSSQVNLITLITNTERRHPRGEATFPQLPRMSGCCGRRLRKAQCTFNSPYRGHY 360
Qy 361 PNIDCTWNIIEVPNNQHVKRFKFFYLBPVGVPAGTCKPDVVEINGEKYGEERSQFVVT 420
Db 361 PNIDCTWNIIEVPNNQHVKRFKFFYLBPVGVPAGTCKPDVVEINGEKYGEERSQFVVT 420
Qy 421 NSNKITVRFHSDOSYTDYGFLEAYLSYDSSDPCPGQFTCRGRCIRKELRCGADCTDH 480
Db 421 NSNKITVRFHSDOSYTDYGFLEAYLSYDSSDPCPGQFTCRGRCIRKELRCGADCTDH 480
Qy 481 SDELNCSGDAGHQPTCKNFKCKPLFWCDSVNDGSDSDGSCCPAQTFRCSNGKCLSK 540
Db 481 SDELNCSGDAGHQPTCKNFKCKPLFWCDSVNDGSDSDGSCCPAQTFRCSNGKCLSK 540
Qy 541 SQQNGKDDCGDSDDEASCPKNNVVTCTKATYRCLNGLCLSKNPECDCGKEDCDGSDDEK 600
Db 541 SQQNGKDDCGDSDDEASCPKNNVVTCTKATYRCLNGLCLSKNPECDCGKEDCDGSDDEK 600
Qy 601 DCDGRLSPTROARVVGTDADGEMPMQVSLHALGGGHI CGASLISPNMLVSAHACYID 660
Db 601 DCDGRLSPTROARVVGTDADGEMPMQVSLHALGGGHI CGASLISPNMLVSAHACYID 660
Qy 661 DRGRYSDPTQWTFALGLHDQSASAPGVQERRLKRIISHPFDFTFDDIALLELEXP 720
Db 661 DRGRYSDPTQWTFALGLHDQSASAPGVQERRLKRIISHPFDFTFDDIALLELEXP 720
Qy 721 AEYSWVRPCLPDASHVFPAGKAIWTTGHTGYGTGALLQKGEIRVINQTTCENL 780
Db 721 AEYSWVRPCLPDASHVFPAGKAIWTTGHTGYGTGALLQKGEIRVINQTTCENL 780
Qy 781 PQQITPRMCCVGLSGVDSQCGSGPLSSVEADGRIFGAGVVSMDGCGAKRKPQVYT 840
Db 781 PQQITPRMCCVGLSGVDSQCGSGPLSSVEADGRIFGAGVVSMDGCGAKRKPQVYT 840
Qy 841 RLPLFRDMIKENTGV 855
Db 841 RLPLFRDMIKENTGV 855
RESULT 25
AD128861
ID AD128861 standard; protein; 855 AA.
XX AC AD128861;
XX

DT 22-APR-2004 (first entry)
XX Human matrixinase (MTSP1) serine protease.
XX Human, matrixinase; MTSP1; CVPSP16; cytostatic; gene therapy; tumour;
KM marker; serine protease.
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FT Domain 615..855
/note="Protease domain"
XX
XX WO2004005471-A2.
XX 15-JAN-2004.
XX
XX 01-JUL-2003; 2003WO-US020959.
XX
XX 02-JUL-2002; 2002US-0394347P.
XX
XX (DEND-) DENDREON SAN DIEGO LLC.
XX
XX Madison EL, Ong EO, Yeh J;
XX
XX MPI; 2004-099379/10.
XX N-PSDB; ADI28860.
XX
XX
XX New serine protease 16 polypeptides, useful for diagnosing, preventing or
PT treating cancer (e.g. breast cancer) or for identifying compounds that
PT may be used for modulating the polypeptides.
XX
XX Disclosure; SEQ ID NO 2; 205pp; English.
XX
XX The present sequence is the protein sequence of human MTSP1 (matrixinase),
CC a serine protease that is expressed in epithelial cancer and normal
CC tissue. In an example from the invention, the protein sequence of the
CC protease domain of MTSP1 was used to search a human genome database. A
CC protease was identified and designated serine protease 16 or CVPSP16.
CC CVPSP16 cDNA ADI28863 was subsequently cloned from a human liver cDNA
CC library by PCR. CVPSP16 polypeptides exhibit protease activity as a single
CC chain or as a multi-chain form. Methods are provided for identifying
CC compounds that modulate the protease activity. CVPSP16 polypeptides also
CC serve as tumour markers.
XX
XX Sequence 855 AA:
SQ
Query Match 99.4%; Score 4653; DB 8; Length 855;
Best Local Similarity 99.5%; Pred. No. 2.4e-307;
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 MGSDRAKGGGPGKDFAGLKYNSRHKNGLEGEVFLPVNNVKYKKGPRGWVLA 60
DB 1 MGSDRAKGGGPGKDFAGLKYNSRHKNGLEGEVFLPVNNVKYKKGPRGWVLA 60
QY 61 VLIGLLVLLIGLGLVHLYRDVRYOVKXNGYRITNENFVADYENSNSSTEFVSLASKV 120
DB 61 VLIGLLVLLIGLGLVHLYRDVRYOVKXNGYRITNENFVADYENSNSSTEFVSLASKV 120
QY 121 KDALKLILYSGVPLGPRHKSATVAESEGVIATYNEEFSPQLVYEAEARVMAEERVM 180
DB 121 KDALKLILYSGVPLGPRHKSATVAESEGVIATYNEEFSPQLVYEAEARVMAEERVM 180
QY 121 KDALKLILYSGVPLGPRHKSATVAESEGVIATYNEEFSPQLVYEAEARVMAEERVM 180
DB 121 KDALKLILYSGVPLGPRHKSATVAESEGVIATYNEEFSPQLVYEAEARVMAEERVM 180
QY 181 LPPRARSLKSFVVTSSVVAAPFDSKTVOITODNSCSFGILHARGVLAEMFTTGGFPDSEYPA 240
DB 181 LPPRARSLKSFVVTSSVVAAPFDSKTVOITODNSCSFGILHARGVLAEMFTTGGFPDSEYPA 240
QY 241 HARCQMALRGDADSVLSLTFRSFDLASCDERGSDLVTVTNTLSPEMBHALVOLCGTYPSP 300
DB 241 HARCQMALRGDADSVLSLTFRSFDLASCDERGSDLVTVTNTLSPEMBHALVOLCGTYPSP 300
QY 301 YNLFHSSQNVLLITLITNTERHRRPGFEATTFQLPRMSSCGGLRKAQGTFSNPIYBGHY 360
DB 301 YNLFHSSQNVLLITLITNTERHRRPGFEATTFQLPRMSSCGGLRKAQGTFSNPIYBGHY 360

DB 301 YNLFHSSQNVLLITLITNTERHRRPGFEATTFQLPRMSSCGGLRKAQGTFSNPIYBGHY 360
QY 361 PNIDCTWNIIEVBNNOHVKVRFPKFFYLLEBGVAGTCEPKDYVEINSEKYGESRQFVTS 420
DB 361 PNIDCTWNIIEVBNNOHVKVRFPKFFYLLEBGVAGTCEPKDYVEINSEKYGESRQFVTS 420
QY 421 NSNKITVRHSDQSYDTGFLAELYSYDSDPCPGQFTCTGACIRKELRCGMADCTDH 480
DB 421 NSNKITVRHSDQSYDTGFLAELYSYDSDPCPGQFTCTGACIRKELRCGMADCTDH 480
QY 481 SDELNCSCAAGHOFCTCKNFKCKPLFWVCDSDVNDGNSDQGSCEPAQTFRCSNGKCLSK 540
DB 481 SDELNCSCAAGHOFCTCKNFKCKPLFWVCDSDVNDGNSDQGSCEPAQTFRCSNGKCLSK 540
QY 541 SQQCNKGDGSDSDASCPKXNVVCTKHTYRCLNGLCLSKGNPRDGEDSDSDSD 600
DB 541 SQQCNKGDGSDSDASCPKXNVVCTKHTYRCLNGLCLSKGNPRDGEDSDSDSD 600
QY 601 DCDGGLRSFTQARVVGGTDADEGEWPMQVSLHALQGHICGASLISPNWLVSAAHCYID 660
DB 601 DCDGGLRSFTQARVVGGTDADEGEWPMQVSLHALQGHICGASLISPNWLVSAAHCYID 660
QY 661 DRGFRYSPTQWTAFLGLHDQSORSA PGVERLKRILISHPFNDFTFDIALLELEXP 720
DB 661 DRGFRYSPTQWTAFLGLHDQSORSA PGVERLKRILISHPFNDFTFDIALLELEXP 720
QY 721 AEYSMWVRPICLPDASHVPAGKAIWVTGHTQYGGTGLILQKEIRVINOCTENLL 780
DB 721 AEYSMWVRPICLPDASHVPAGKAIWVTGHTQYGGTGLILQKEIRVINOCTENLL 780
QY 781 PQOITPRMNCVGLSGVDSCQDSSGPISSVADGRIRAGVYVSGDGCAGNKKGYVT 840
DB 781 PQOITPRMNCVGLSGVDSCQDSSGPISSVADGRIRAGVYVSGDGCAGNKKGYVT 840
QY 841 RLPLFRDWIKENTGV 855
DB 841 RLPLFRDWIKENTGV 855
RESULT 26
ADJ46895
ID ADJ46895 standard; protein; 855 AA.
XX
AC ADJ46895;
XX
DT 06-MAY-2004 (first entry)
XX
XX Human transmembrane serine protease (MTSP) polypeptide #1.
XX
XX Human; transmembrane serine protease; MTSP; cell surface protease;
KM plasmin; prostate specific antigen; PSA; proliferative disease;
KM cell-surface protease-associated disease; autoimmune disease;
KM inflammatory disease; infectious disease; endocrine disease; cancer;
KM ocular disorder; cardiovascular disorder; chronic inflammatory disease;
KM wound; circulatory disorder; dermatological disorder; testecrosis;
KM rheumatoid arthritis; psoriasis; diabetic retinopathy;
KM laser surgery scarring; glaucoma filtering surgery scarring;
KM macular degeneration; CRST syndrome; bacterial infection; viral disease;
KM solid neoplasm; vascular tumour; lung; colon; prostate; melanoma;
KM Kaposi's sarcoma; enzyme.
XX
XX Homo sapiens.
XX
XX US2004001801-A1.
XX
XX 01-JAN-2004.
XX
XX 23-MAY-2002; 2002US-00156214.
XX
XX 23-MAY-2002; 2002US-00156214.
XX
XX (CORV-) CORVAS INT INC.
XX

PI Madison EL, Sempke JE, Vlasuk GP, Kemp SJ, Komandla M, Siew DV;
XX WPI, 2004-190126/18.
DR N-PSDB; ADU46894.
XX Conjugate useful for treating e.g. cancer, cell-surface protease-
PT associated diseases, comprising a peptidic substrate or nucleic acid
PT substrate linked to a therapeutic agent through a linker.
XX Claim 9; SEQ ID NO 2; 361bp; English.
XX The invention relates to a conjugate comprising a therapeutic agent and a
CC peptidic substrate or nucleic acid substrate linked to the agent
CC optionally through a linker or peptidic linker, where the peptidic
CC substrate is proteolytically cleaved by a cell surface protease or a
CC soluble, released or shed form conjugate to liberate the agent and the
CC conjugate is not substantially cleaved by plasmin or prostate specific
CC antigen (PSA). The conjugate is useful for treating a disease, which
CC involves administering a conjugate to a subject, where the disease is
CC preferably a proliferative diseases or a cell-surface protease-associated
CC disease. The diseases include autoimmune diseases, inflammatory diseases,
CC infectious diseases and endocrine diseases. The conjugate is useful for
CC treating a cell-surface protease-associated disease, which involves
CC administering a conjugate comprising an agent and a peptidic substrate to
CC a subject exhibiting symptoms of a cell-surface protease-associated
CC disorder, where the disease is selected from cancer, ocular disorders,
CC cardiovascular disorders, chronic inflammatory diseases, wounds,
CC circulatory disorders, dermatological disorders, restenosis, rheumatoid
CC arthritis, psoriasis, diabetic retinopathies, scarring from laser
CC surgery, scarring from glaucoma filtering surgery, macular degeneration,
CC CRST syndrome, bacterial infections, viral diseases, solid neoplasms and
CC vascular tumors such as lung cancer, colon cancer, prostate cancer,
CC melanoma and Kaposi's sarcoma. This sequence represents a transmembrane
CC serine protease (Misp) polypeptide of the invention.
XX
SQ Sequence 855 AA;
Query March 99.4%; Score 4653; DB 8; Length 855;
Beet Local Similarity 99.5%; Pred. No. 2.4e-307;
Matches 851; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 MGSDRARKGGGPGDFGAGLKYNSRHEKNGLEGEVEFLPVNNVYKKGKPGHVVVLA 60
Db 1 MGSDRARKGGGPGDFGAGLKYNSRHEKNGLEGEVEFLPVNNVYKKGKPGHVVVLA 60
QY 61 VLIGLLVLLGIGFLVHLYQYRDVAVQKVKNGKYNITENFVDAYENSSTEFVSLASKV 120
Db 61 VLIGLLVLLGIGFLVHLYQYRDVAVQKVKNGKYNITENFVDAYENSSTEFVSLASKV 120
QY 121 KDAIKLYSGVPLGPRYHKEASVTAFAFSGSVIAYYWSFSLPHLYVEAEVMAEERVVM 180
Db 121 KDAIKLYSGVPLGPRYHKEASVTAFAFSGSVIAYYWSFSLPHLYVEAEVMAEERVVM 180
QY 181 LPPPARSLKSFVYTSVVAFPDTSKTVQRTQNSCSFGJAHAGVELMRETTGFPDPSPYPA 240
Db 181 LPPPARSLKSFVYTSVVAFPDTSKTVQRTQNSCSFGJAHAGVELMRETTGFPDPSPYPA 240
QY 241 HARCQMALRGDADSVLSTFRSPDLASCDERGSIDLVTYNTLSPMEPHALVOLCGTYPSS 300
Db 241 HARCQMALRGDADSVLSTFRSPDLASCDERGSIDLVTYNTLSPMEPHALVOLCGTYPSS 300
QY 301 YNLTFHSSQNVLLTLITNTERRHGFEATFQQLPRMSSCGRLRKAGCTNSPPYRGHY 360
Db 301 YNLTFHSSQNVLLTLITNTERRHGFEATFQQLPRMSSCGRLRKAGCTNSPPYRGHY 360
QY 361 PPNIDCTNINIEVNNQHVAFKFFYLLEPGVPAATCKDVEYINGEKYCGERSQFVYTS 420
Db 361 PPNIDCTNINIEVNNQHVAFKFFYLLEPGVPAATCKDVEYINGEKYCGERSQFVYTS 420
QY 421 NSNKITVRFHSDQSTYDTGFLAEVLYSYDSGPCPGQFTCRGRCTIRKELRCDGADCTDH 480
Db 421 NSNKITVRFHSDQSTYDTGFLAEVLYSYDSGPCPGQFTCRGRCTIRKELRCDGADCTDH 480

QY 481 SDELINSCDAGHQTCKNKECKPLFWVCDVNDGNSDEQSCSPAQTRFRCNSGKCLSK 540
Db 481 SDELINSCDAGHQTCKNKECKPLFWVCDVNDGNSDEQSCSPAQTRFRCNSGKCLSK 540
QY 541 SQQNGKXDCGDSDEASCPKRVNVVCTKTYRCLNGLCLSKNGPECDGKEDCSGSDXK 600
Db 541 SQQNGKXDCGDSDEASCPKRVNVVCTKTYRCLNGLCLSKNGPECDGKEDCSGSDXK 600
QY 601 DCDGGLASFTRQARVVGCTADDEGEPMQVSHLAGGSHCGASLISPMNLVSAHGYID 660
Db 601 DCDGGLASFTRQARVVGCTADDEGEPMQVSHLAGGSHCGASLISPMNLVSAHGYID 660
QY 661 DRGRYSDPTQWTFALGHDQSORAPGVERLRKRIISHPFNDFPDYDIALLEDEKP 720
Db 661 DRGRYSDPTQWTFALGHDQSORAPGVERLRKRIISHPFNDFPDYDIALLEDEKP 720
QY 721 AEYSMWBPICLPDASHVFPAGKAIWVTGWGHTQYGGTGALLQKGEIRVINQTCENLL 780
Db 721 AEYSMWBPICLPDASHVFPAGKAIWVTGWGHTQYGGTGALLQKGEIRVINQTCENLL 780
QY 781 PQQITPRMVCVPLSGVDSQGGSGPLSSVEADGRIFGAGVWSKDGCAGRKPGVYT 840
Db 781 PQQITPRMVCVPLSGVDSQGGSGPLSSVEADGRIFGAGVWSKDGCAGRKPGVYT 840
QY 841 RLPLFRDMIKENTGV 855
Db 841 RLPLFRDMIKENTGV 855
RESULT 27
AD055145
ID AD055145 standard; protein: 853 AA.
AC AD055145;
DT 15-JUN-2004 (first entry)
XX
DE Protein #47 with increased gene expression in renal cell carcinoma.
XX
KW cyrostatic; gene therapy; differential expression; renal cell carcinoma;
KW clear cell RCC; papillary RCC; chromophobe/oncocytoma RCC;
KW sarcomatoid RCC; RCC; Wilms' tumor; gene expression; kidney cancer;
KW diagnostic marker; cancer.
OS Homo sapiens.
PN MO2004032842-A2.
PD 22-APR-2004.
PP 06-OCT-2003; 2003WO-US031476.
PR 04-OCT-2002; 2002US-0415775P.
PA (VAND-) VAN ANDEL INST.
PI Teh BT, Takahashi M;
XX WPI, 2004-340789/31.
DR N-PSDB; AD054959.
XX
PT New nucleic acid and polypeptide compositions, useful in the field of
PT molecular biology and medicine, in particular for gene expression
PT profiling, identifying diagnostic markers, and treating certain types of
PT kidney cancer.
XX
PS Example IV; SEQ ID NO 242; 53bp; English.
XX
CC The invention relates to novel genes that are differentially expressed in
CC sub-types of renal cell carcinomas and methods of detecting them using
CC nucleic acids and probes. The nucleic acid probes hybridize with part or
CC all of a coding sequence that is overexpressed in clear cell renal cell
CC carcinoma (CC-RCC), papillary RCC, chromophobe/oncocytoma RCC,

CC sarcomatoid RCC, RCC, or Wilms' tumors, which overexpression is based on
CC comparison to a baseline value. The methods and compositions of the
CC present invention are useful in the field of molecular biology and
CC medicine, in particular for gene expression profiling of certain types of
CC kidney cancer, in identifying diagnostic markers, and treating such
CC cancer patients. This sequence corresponds to the protein encoded by a
CC gene with increased expression in CC-RCC.

XX Sequence 853 AA:

Query Match 98.8%; Score 4625; DB 8; Length 853;

Best Local Similarity 99.4%; Pred. No. 1,9e-305;

Matches 850; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

```

QY 1 MGSPRRKGGGGPPDPFAGLKYNRHRKNGLEGEVFLPVNNVKKYKPGKPGWVLLAA 60
Db 1 MGSPRRKGGGGPPDPFAGLKYNRHRKNGLEGEVFLPVNNVKKYKPGKPGWVLLAA 60
QY 61 VLIIGLLVLIIGLIGFLVHLYRDVRYQKYNKGYMRITNENFVADYENSNSTEFVSLASKV 120
Db 61 VLIIGLLVLIIGLIGFLVHLYRDVRYQKYNKGYMRITNENFVADYENSNSTEFVSLASKV 120
QY 121 KDALKLLYSGVPLGPHKKSATVAFSEGSYIATYNEFSFPGHLYVEAEKRWAEKRVVM 180
Db 121 KDALKLLYSGVPLGPHKKSATVAFSEGSYIATYNEFSFPGHLYVEAEKRWAEKRVVM 180
QY 181 LPPRRSLKSFVTVSVVAFPTDSKTVQRTODNSCFGHARGVLYMEFTTPGFPDSDPYPA 240
Db 181 LPPRRSLKSFVTVSVVAFPTDSKTVQRTODNSCFGHARGVLYMEFTTPGFPDSDPYPA 240
QY 241 HARCQMLRGDADSVLSLTFRSFDLASCDERGSDLVTVYNTLSPEMHALVOLCGTYPSP 300
Db 240 HA-CQMLRGDADSVLSLTFRSFDLASCDERGSDLVTVYNTLSPEMHALVOLCGTYPSP 300
QY 301 YNLFHSSQNVLLITLTNTERRHHPGEATFPQLPRMSSCGGRARKAOGTFNSPYRGHY 360
Db 301 YNLFHSSQNVLLITLTNTERRHHPGEATFPQLPRMSSCGGRARKAOGTFNSPYRGHY 360
QY 361 PPNIDCTWNIIEVPNNQHVAFKFFYLLRPGVAGTCKPKQVEINGEKYCGERSQFVVT 420
Db 361 PPNIDCTWNIIEVPNNQHVAFKFFYLLRPGVAGTCKPKQVEINGEKYCGERSQFVVT 420
QY 421 NSNKITVRFHSDQSYDTGTFLAEYLSYDSDPCPGQFTCRTGRCIRKELRCGNADCTDH 480
Db 421 NSNKITVRFHSDQSYDTGTFLAEYLSYDSDPCPGQFTCRTGRCIRKELRCGNADCTDH 480
QY 481 SDELNCSCDAGHOTCTCKNPKCKPLFWICDSVNDGDSDEGSCCPAQTFRCSNGKCLSK 540
Db 481 SDELNCSCDAGHOTCTCKNPKCKPLFWICDSVNDGDSDEGSCCPAQTFRCSNGKCLSK 540
QY 541 SQQNGSDDCGSDGDEASCPKVNVTCTKHITYRCIANGLCISKGNPECDGKEDCSDGSD 600
Db 541 SQQNGSDDCGSDGDEASCPKVNVTCTKHITYRCIANGLCISKGNPECDGKEDCSDGSD 600
QY 601 DCDGGLRSFTRQARVVGTDADDEGEWQVLSLHLAGQSHICGASLISPNMLVSAHICYID 660
Db 601 DCDGGLRSFTRQARVVGTDADDEGEWQVLSLHLAGQSHICGASLISPNMLVSAHICYID 660
QY 661 DRGRVSDPQWMTAFELIHDSQSRSAQVQERRIKRIISHPFNDPFEDVDIALLELEKP 720
Db 661 DRGRVSDPQWMTAFELIHDSQSRSAQVQERRIKRIISHPFNDPFEDVDIALLELEKP 720
QY 721 AEYSMTVPICLPASHVFPAGKAIWYGMGHTQYGGTALILQKGIKRVINOTTCEML 780
Db 721 AEYSMTVPICLPASHVFPAGKAIWYGMGHTQYGGTALILQKGIKRVINOTTCEML 780
QY 781 PQQITPRMWCVGLISGVDSQGDSSGGLSSVEADGRIFGAGVSVWSDGCAGRNKPQYTT 840
Db 781 PQQITPRMWCVGLISGVDSQGDSSGGLSSVEADGRIFGAGVSVWSDGCAGRNKPQYTT 840
QY 841 RLPLFRDWIKENTGV 855
Db 841 RLPLFRDWIKENTGV 855
QY 853 RLPLFRDWIKENTGV 853
Db 853 RLPLFRDWIKENTGV 853

```

RESULT 28

AAM25628
ID AAM25628 standard; protein; 851 AA.

XX AAM25628;

DT 16-OCT-2001 (first entry)

DE Human protein sequence SEQ ID NO:1143.

XX Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
KM antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;
KM antibacterial; endocrine; cardiant; central nervous system; virucide;
KM anti-HIV; fungicide; antimutagen; cardiovascular; antinaemic; anaemia;
KM antiaggregant; haemostatic; vulnery; antilucer; osteopathic; eczema;
KM dermatological; anti allergic; antistatic; antidiabetic; cyostatic;
KM neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
KM immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
KM antiaphyllactic; rheumatoid arthritis; septic shock; pancreatitis;
KM cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
KM genetic disease; haematopoietic disorder; platelet disorder; asthma;
KM thrombocytopenia; osteoporosis; severe combined immunodeficiency;
KM allergic rhinitis; diabetes; multiple sclerosis; depression;
KM Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
KM neurological disorder.

OS Homo sapiens.

PN WO200153455-A2.

PD 26-JUL-2001.

PF 22-DEC-2000; 2000MO-US035017.

PR 23-DEC-1999; 99US-00471275.

PR 21-JAN-2000; 2000US-00488725.

PR 25-APR-2000; 2000US-00552317.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Drmanac RT;

DR WPI; 2001-457603/49.

PT N-PSDB; AAM99569.

PS Isolated human polynucleotides encoding polypeptides, useful for the

XX treatment and diagnosis of e.g. cancer, ulcers and HIV infection.

XX Claim 20; Page 237; 1217pp; English.

XX AAM99166 to AAM99994 encode the human proteins given in AAM25225 to

XX AAM25863. The proteins can have activities based on the tissues and cells

XX they are expressed in, such as: antiinflammatory; antirheumatic;

XX antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;

XX central nervous system; virucide; anti-HIV; fungicide; antimutagen;

XX cardiovascular; antinaemic; antiaggregant; haemostatic; vulnery;

XX antilucer; osteopathic; dermatological; anti allergic; antistatic;

XX antidiabetic; cyostatic; neuroprotective; antidepressant; nootropic;

XX antiparkinsonian; and immunostimulant. The proteins and polynucleotides

XX encoding them can be used in gene therapy, antisense therapy and vaccine

XX production. The proteins and polynucleotides are useful for screening for

XX agonists or antagonists of a protein and for the treatment and diagnosis

XX of disorders associated with the activity of a protein e.g. inflammation,

XX rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,

XX neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal

XX infections, autoimmunity, genetic diseases, haematopoietic disorders,

XX anaemia, platelet disorders, thrombocytopenia, wounds, burns, ulcers,

XX osteoporosis, severe combined immunodeficiency, eczema, allergic

XX rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,

XX Alzheimer's disease, Parkinson's disease, neurodegenerative and

XX neurological disorders

XX Sequence 851 AA:
SQ Query Match 98.5%; Score 4614; DB 4; Length 851;
Best Local Similarity 99.3%; Pred. No. 1.1e-304;
Matches 843; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 7 RKGGGGKPDPAAGLKYNRRHKNVNGLEGGVFLPVNNYKTKKPKGPMVLAIVLIGLL 66
DB 3 RKGGGGPDPAAGLKYNRRHKNVNGLEGGVFLPVNNYKTKKPKGPMVLAIVLIGLL 62
QY 67 LVLLIGFLVHMLQYRDVAVQKXNGVYRINENFVDAEYNSNTEFSLASKYDALKL 126
DB 63 LVLLIGFLVHMLQYRDVAVQKXNGVYRINENFVDAEYNSNTEFSLASKYDALKL 122
QY 127 LYSVGPFLGPKYHKSAAVAFSEGSVIAVYMSFSPQHLVEABERVMAEERVNLPPRAR 186
DB 123 LYSVGPFLGPKYHKSAAVAFSEGSVIAVYMSFSPQHLVEABERVMAEERVNLPPRAR 182
QY 187 SLKSFVVTSVVAFPTDSTKVORTODNSGSGFLHARGVELMFTTPGPPDSYPHARQW 246
DB 183 SLKSFVVTSVVAFPTDSTKVORTODNSGSGFLHARGVELMFTTPGPPDSYPHARQW 242
QY 247 ALRGDADSVLSLTFRSFPLASCDERGSDDLVTYNTLSPEMHALVOLGCTTPSPSYNLTFH 306
DB 243 ALRGDADSVLSLTFRSFPLASCDERGRHLVTYNTLSPEMHALVOLGCTTPSPSYNLTFH 302
QY 307 SSQNVLLITLITNTERRHPPGEATFPQLPRMSSCGGLRKAQGTFSNPPYPGHYPPNIDC 366
DB 303 SSQNVLLITLITNTERRHPPGEATFPQLPRMSSCGGLRKAQGTFSNPPYPGHYPPNIDC 362
QY 367 TWINIEVNNQVKKRPFKFFYLLEPGVPAGTCKPYVEINERKXCGERSQFVYTSNSKIT 426
DB 363 TWINIEVNNQVKKRPFKFFYLLEPGVPAGTCKPYVEINERKXCGERSQFVYTSNSKIT 422
QY 427 VRFHSDQSYDTGTGLAEYLSYSDSDPCPGQFTCRGRCIRRELKCDGADCTDSDELNC 486
DB 423 VRFHSDQSYDTGTGLAEYLSYSDSDPCPGQFTCRGRCIRRELKCDGADCTDSDELNC 482
QY 487 SCDAHQFTCKNKKCPFLFWVCDSDVNDGSDSDGSCSPAQTFRCNSGKCLSKSQQNG 546
DB 483 SCDAHQFTCKNKKCPFLFWVCDSDVNDGSDSDGSCSPAQTFRCNSGKCLSKSQQNG 542
QY 547 KDDCGDSDASCRVNVVTTKITYRCLNGLCLSKNPEBDGKXDCSDGSDDEKDCDGL 606
DB 543 KDDCGDSDASCRVNVVTTKITYRCLNGLCLSKNPEBDGKXDCSDGSDDEKDCDGL 602
QY 607 RSPTRQARVVGTDADGEMPMOYSLHALGQGHICGASLISPNMLVSAHACYIDRGRY 666
DB 603 RSPTRQARVVGTDADGEMPMOYSLHALGQGHICGASLISPNMLVSAHACYIDRGRY 662
QY 667 SDPTQMTAFGLHDSQSRAPGVORRLKRIISHPFNDFTFDYDIALLEKPAEYSSM 726
DB 663 SDPTQMTAFGLHDSQSRAPGVORRLKRIISHPFNDFTFDYDIALLEKPAEYSSM 722
QY 727 VRPITCLPDASHVPFAGKAIWTTGNGHTQYGGTGALLQKGEIRYINOTTGNNLPQQTTP 786
DB 723 VRPITCLPDASHVPFAGKAIWTTGNGHTQYGGTGALLQKGEIRYINOTTGNNLPQQTTP 782
QY 787 RMMCVCGLFSGVDSGQSGGGLSSVEADGRIFGAGVSVWMDGCGARRKPGVYTRLPFLR 846
DB 783 RMMCVCGLFSGVDSGQSGGGLSSVEADGRIFGAGVSVWMDGCGARRKPGVYTRLPFLR 842
QY 847 DWIKENTGV 855
DB 843 DWIKENTGV 851

RESULT 29
ABBI1428 standard; peptide; 851 AA.
XX
AC ABBI1428;

XX 11-JAN-2002 (first entry)
DE Human membrane-type Ser kinase homologue, SEQ ID NO:1798.
XX Human; cytokine; cell proliferation; cell differentiation; growth factor;
XX haematopoietic regulation; tissue growth; immunomodulator; actinin;
XX inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
XX proliferation; metastasis; cancer; tumour; haematopoietic disorder;
XX myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
XX chronic inflammatory condition; proliferative retinopathy;
XX atherosclerosis; coronary heart disease; arterial ischaemia;
XX bone disorder; osteoporosis; vascular growth disorder;
XX tissue regeneration; wound healing; infection; immune disorder;
XX cell culture; drug screening; gene therapy; antiinflammatory;
XX antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
XX cytoskeletal; osteopathic; vasotrophic; cardiant; virucide; antibacterial;
XX antifungal; vulnery; antitumor.
XX Homo sapiens.
XX WO200157188-A2.
XX 09-AUG-2001.
XX 05-FEB-2001; 2001MO-US003800.
XX 03-FEB-2000; 2000US-00496914.
XX 27-APR-2000; 2000US-00560875.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Drmanac RT;
XX WPI; 2001-457740/49.
XX N-PSDB; ABA08672.
XX Human proteins and DNA encoding sequences useful for preventing, treating
XX PT or ameliorating a medical condition in a mammalian subject e.g. arthritis
XX PT and cancer.
XX Claim 20; Page 188; 1963pp; English.
XX
XX Sequences ABBI0981-ABBI2330 represent 1350 novel human polypeptides, and
XX sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
XX invention also relates to vectors and recombinant host cells comprising a
XX nucleotide of the invention, methods of producing the novel polypeptides,
XX antibodies against the polypeptides, methods of detecting the nucleotides
XX or polypeptides in a sample, and methods of identifying compounds which
XX bind to polypeptides of the invention. Although novel, many of the
XX polypeptides of the invention have homology to known proteins, thereby
XX giving an insight into their probable biological activities, and hence
XX potential therapeutic applications. The polypeptides of the invention may
XX have various activities, including cytokine, cell proliferation or cell
XX differentiation activities; stem cell growth factor activity;
XX haematopoietic regulatory activity; tissue growth activity;
XX immunomodulatory activity; actinin- or inhibin-related activities;
XX chemotactic or chemokinetic activities; haemostatic, thrombotic or
XX thrombolytic activities; receptor or ligand activities; or may be
XX involved in oncogenesis, cancer cell proliferation or metastasis.
XX Depending on their biological activities, polypeptides and nucleotides of
XX the invention are useful for preventing, treating or ameliorating medical
XX conditions, e.g., by protein or gene therapy. Such conditions include
XX cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell
XX disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
XX proliferative retinopathy, atherosclerosis, coronary heart disease,
XX arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
XX vascular growth. Polypeptides involved with tissue regeneration and
XX repair (or nucleic acids encoding them) may be used to promote wound
XX healing (e.g., of burns, incisions and ulcers), while those with
XX immunomodulatory activities may be used in the treatment of viral,
XX bacterial and fungal infections in addition to immune disorders.
XX Polypeptides with growth factor activity may be used in cell cultures to

CC promote cell growth. For example, such polypeptides may be used to
 CC manipulate stem cells in culture to give rise to neuroepithelial cells
 CC that can be used to augment or replace cells damaged by illness,
 CC autoimmune disease or accidental damage. The polypeptides and nucleotides
 CC may also be used in the diagnosis of the above conditions, and in drug
 CC screening techniques. The present sequence represents a novel human
 CC polypeptide of the invention
 CC
 XX

XX Sequence 851 AA;

Query Match 98.5%; Score 4614; DB 4; Length 851;

Best Local Similarity 99.3%; Pred. No. 1.1e-304;

Matches 843; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 7 RKGGGGFKDYGAGLKYNSRHEKXVNGLEGEVFLPVNNVKYKXKPGRWVLAAILGLL 66
 DB 3 RKGGGGFKDYGAGLKYNSRHEKXVNGLEGEVFLPVNNVKYKXKPGRWVLAAILGLL 62
 QY 67 LVTLIGFLVWHLQYRDVRYQKXNGYMRITNENFVDAYENSNSTEFVSLASKYKDALKL 126
 DB 63 LVTLIGFLVWHLQYRDVRYQKXNGYMRITNENFVDAYENSNSTEFVSLASKYKDALKL 122
 QY 127 LVSGVPELPGYHKEASATAFSEGSVIAVYVSEFIPQHLVEAERVAEERVWMLPPRAR 186
 DB 123 LVSGVPELPGYHKEASATAFSEGSVIAVYVSEFIPQHLVEAERVAEERVWMLPPRAR 182
 QY 187 SLKSFVTVSVAFPEPDSKTVQRTQDNSCFGLHARGVELMRFTTPGPPDSPYPAHARQW 246
 DB 183 SLKSFVTVSVAFPEPDSKTVQRTQDNSCFGLHARGVELMRFTTPGPPDSPYPAHARQW 242
 QY 247 ALRGDADSVTLSTFRSDLASCDERGDLYTVNTLSPMERHALVOLCGTTPPSYNTLTFH 306
 DB 243 ALRGDADSVTLSTFRSDLASCDERGDLYTVNTLSPMERHALVOLCGTTPPSYNTLTFH 302
 QY 307 SSONVLLITLITNTERHHPGEATFPLPMSSCGRLRAQGFENSPYRPHYPIVDC 366
 DB 303 SSONVLLITLITNTERHHPGEATFPLPMSSCGRLRAQGFENSPYRPHYPIVDC 362
 QY 367 TWINIEVNNQHVKKRFKFFYLLEPGVAGTCPKDYVEINGEKXCGERSQEPVTSNSNKT 426
 DB 363 TWINIEVNNQHVKKRFKFFYLLEPGVAGTCPKDYVEINGEKXCGERSQEPVTSNSNKT 422
 QY 427 VRHHSDDSYDTGTLAETLYSDSDPCPGQFTCTGTCIRKEALRCDMACTHSDSLNC 486
 DB 423 VRHHSDDSYDTGTLAETLYSDSDPCPGQFTCTGTCIRKEALRCDMACTHSDSLNC 482
 QY 487 SCAGHOFCTCKNKKCKPLFWVCDSDVNDGNSDDEGSCPAQPRGCSNGKCLSKSQOQNG 546
 DB 483 SCAGHOFCTCKNKKCKPLFWVCDSDVNDGNSDDEGSCPAQPRGCSNGKCLSKSQOQNG 542
 QY 547 KDDCGDSDBASCPKXNVVTCTKTYRCLNGLCLSKGNPECDGKEDCDSDDEKDCDGL 606
 DB 543 KDDCGDSDBASCPKXNVVTCTKTYRCLNGLCLSKGNPECDGKEDCDSDDEKDCDGL 602
 QY 607 RSFTRQARVVGCTDADGEMPMQVSLALGCGHICGASLISPMWLVAHAACIYIDRGRY 666
 DB 603 RSFTRQARVVGCTDADGEMPMQVSLALGCGHICGASLISPMWLVAHAACIYIDRGRY 662
 QY 667 SDPTQWAFPLGLHDQSORSAFGVQERLKRILISHPFNDTFPVDIALLELEKPAEYSSM 726
 DB 663 SDPTQWAFPLGLHDQSORSAFGVQERLKRILISHPFNDTFPVDIALLELEKPAEYSSM 722
 QY 727 VRPILCPDASHVPBAGKAIWVTGHTOYGGTGALLQKBEIRVINTTCEMLPQGITP 786
 DB 723 VRPILCPDASHVPBAGKAIWVTGHTOYGGTGALLQKBEIRVINTTCEMLPQGITP 782
 QY 787 RMCVCGLSGCVDSCCGDSGGPLSSVADGRIFGAGVVSNGDGCAGNNKGVYTRLPFLR 846
 DB 783 RMCVCGLSGCVDSCCGDSGGPLSSVADGRIFGAGVVSNGDGCAGNNKGVYTRLPFLR 842
 QY 847 DWIKENTGV 855
 DB 843 DWIKENTGV 851

RESULT 30

ABG21442

ID ABG21442 standard; protein; 932 AA.

XX ABG21442;

XX 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #21433.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;

XX food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

XX NC0200175067-A2.

XX 11-OCT-2001.

PF 30-MAR-2001; 2001MO-US008631.

XX 31-MAR-2000; 2000US-00540217.

PR 23-AUG-2000; 2000US-00649167.

XX (HYSE-) HYSEQ INC.

PA Dmanac RT, Liu C, Tang YT;

XX WPI, 2001-639362/73.

DR N-PSDB; AAS85629.

PT New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess

PT biodiversity.

XX Claim 20; SEQ ID NO 51801; 103bp; English.

PS The invention relates to isolated polynucleotide (I) and polypeptide (II)

XX sequences. (I) is useful as hybridisation probes, polymerase chain

XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,

XX and in recombinant production of (II). The polynucleotides are also used

XX in diagnostics as expressed sequence tags for identifying expressed

XX genes. (I) is useful in gene therapy techniques to restore normal

XX activity of (II) or to treat disease states involving (II). (II) is

XX useful for generating antibodies against it, detecting or quantitating a

XX polypeptide in tissue, as molecular weight markers and as a food

XX supplement. (II) and its binding partners are useful in medical imaging

XX of sites expressing (II). (I) and (II) are useful for treating disorders

XX involving aberrant protein expression or biological activity. The

XX polypeptide and polynucleotide sequences have applications in

XX diagnostics, forensics, gene mapping, identification of mutations

XX responsible for genetic disorders or other traits to assess biodiversity

XX and to produce other types of data and products dependent on DNA and

XX amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic

XX patent did not appear in the printed specification, but was obtained in

XX electronic format directly from WIPO at

XX ftp.wipo.int/pub/published_pct_sequences

XX

XX Sequence 932 AA;

XX

Query Match 91.9%; Score 4302; DB 4; Length 932;

Best Local Similarity 95.8%; Pred. No. 1.9e-283;

Matches 796; Conservative 7; Mismatches 24; Indels 4; Gaps 4;

QY 27 EKVNGLEBGEVFLPVNNVKKEGKGPGRWVLAAILIGLLVLLIGLFLVWHLQYRDVRY 86
 DB 100 QKVNGLEBGEVFLPVNNVKKEGKGPGRWVLAAILIGLLVLLIGLFLVWHLQYRDVRY 159
 QY 87 QKVNGYMRITNENFVDAYENSNSTEFVSLASKYKDALKLIVSGVPELPGYHKEASATAF 146

Db 160 QKVNGVIRINENFVDAYENSNTSEFVSLASKVDALAKLYSGVPLGPHKESAATAF 219
QY 147 SEGSVIAYYMEFSI PQHLVEAEERVMAEERVVMLPPPARSLKSFVTSVVAFPDSTKV 206
Db 220 SEGSVIAYYMEFSI PQHLVEAEERVMAEERVVMLPPPARSLKSFVTSVVAFPDSTKV 279
QY 207 QRTODNSCSFGLHARGVELMRFTTPGFPDSDPYPAHARCOMALRGDADSVLSLFRSPDLA 266
Db 280 QRTODNSCSFGLHARGVELMRFTTPGFPDSDPYPAHARCOMALRGDADSVLSLFRSPDLA 339
QY 267 SCDERGSDLVTVYNTLSPEMPEHALVOLCGTYPSPSYNLTFHSSQNVLLITLTNTERRHFG 326
Db 340 SCDERGRILVTVYNTLSPEMPEHALVOLCGTYPSPSYNLTFHSSQNVLLITLTNTERRHFG 399
QY 327 FEATFPQLPRMSSCGGRLRKAQGTFNSPYYPGHYRP-NIDCTWNI EVPNNOHVKKRFKF 385
Db 400 FEATFPQLPRMSSCGGRLRKAQGTFNSPYYPGHYRPQHXSTWNI EVPNNOHVKKRFKF 459
QY 386 YLLERGVAGTCPPDYVEINEKCYGERSQFVVTNSNKTIVRFHSDQSYTDGFLAEXL 445
Db 460 YLLERGVAGTCPPDYVEINEKCYGERSQFVVTNSNKTIVRFHSDQSYTDGFLAEXL 519
QY 446 SYDSSDPGCGQFTCRTR-CIRKELRCDMADCTDHSDELNCSCDAGHQFTCKNFKCKPL 504
Db 520 SYDSSDPGCGQFTCRTRCAGAVIRKELRCDMADCTDHSDELNCSCDAGHQFTCKNFKCKPL 579
QY 505 FMVCDSVNDGSDNSDEQSCCPAQTFRCSNGKCLSKSQCNKGKDDCGDSDASCPKXNV 564
Db 580 FMVCDSVNECGDNSDEQSCCPAQTFRCSNGKCLSKSQCNKGKDDCGDSDASCPKXNV 639
QY 565 VTCTKHTYRCANGCLSKGNPECDGKEDSDGSDKDCDGLRSTTRQARVVGTDADG 624
Db 640 VTCTKHTYRCANGCLSKGNPECDGKEDSDGSDKDCDGLRSTTRQARVVGTDADG 699
QY 625 EPMQVSLHALGQGHICGASLISPMVLVSAHCVIHDRGFRYSDEPTQWT-AFLGLHDQSQ 683
Db 700 EPMQVSLHALGQGHICGASLISPMVLVSAHCVIHDRGFRYSDEPTQWT-AFLGLHDQSQ 759
QY 684 RSAP-GVQERLKKIISHPFNDFTFDYDIALLELEKPAEYSSMVRPICLPDASHVFPAG 742
Db 760 RRPWGCRKRLKRIISHPFNDFTFDYDIALLELEKPAEYSSMVRPICLPDASHVFPAG 819
QY 743 KAIWVTGWHTOYGGTGALLIOKEIRVINOITTCENLLPQOITPRMVCVGLSGVDSQ 802
Db 820 KAIWVTGWHTOYGGTGALLIOKEIRVINOITTCENLLPQOITPRMVCVGLSGVDSQ 879
QY 803 GDSGGLSSVEADGRIFGAGVVMGDCAGRNKPGVYTRLPFRDMIKENTGV 855
Db 880 GDSGGLSSVEADGRIFGAGVVMGDCAGRNKPGVYTRLPFRDMIKENTGV 932

Search completed: November 29, 2004, 08:28:11
Job time : 129.858 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 29, 2004, 08:19:13 : Search time 142.315 Seconds
(without alignments)
3456.743 Million cell updates/sec

Title: US-09-936-333-27

Perfect score: 4683
Sequence: 1 MGSRARAKGGGKPGAGL.....PGVYTRLPFRDKWIKENTGV 855

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database :

1: uniprot_02:.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4659	99.5	855	1	ST14_HUMAN
2	3884	82.9	855	1	SL14_MOUSE
3	3884	82.9	855	2	BAC35125
4	3866	82.6	855	2	O9J117
5	2686	57.4	845	2	O6GR54
6	2660	56.8	845	2	O9DGR1
7	2365	50.5	422	2	O8WVCI
8	1965	42.0	663	2	O6DEV0
9	1114.5	23.8	799	2	O6PF94
10	1114.5	23.8	799	2	AAH57674
11	1114.5	23.8	811	1	TMS6_MOUSE
12	1114	23.7	802	2	AAQ88764
13	1112	23.7	802	2	O6ICCC
14	1112	23.7	802	2	O6ICCC
15	1092	23.3	824	2	O6ICCC
16	1092	23.3	824	2	CAG30332
17	1004.5	21.4	572	2	O7RTY8
18	1004.5	21.4	572	2	O8BIK6
19	1004.5	21.4	572	2	BAD18401
20	724	15.5	1034	1	ENTK_PIG
21	711.5	15.2	855	2	O7Z410
22	711.5	15.2	1059	2	O7Z411
23	709.5	15.2	1035	1	ENTK_BOVIN
24	695	14.8	1042	1	CORI_HUMAN
25	694.5	14.8	1111	2	O8OYN4
26	684	14.6	777	2	O8CAN9
27	679.5	14.5	1019	1	ENTK_HUMAN
28	673.5	14.4	1069	1	ENTK_MOUSE
29	661.5	14.1	1113	1	CORI_MOUSE
30	639.5	13.7	767	2	O9DGR2
31	639	13.6	722	2	O6NUP5

32	639	13.6	722	2	AAH68636	AAH68636 xenopus 1
33	633.5	13.5	680	2	O868H7	O868H7 branchiosteo
34	616.5	13.2	680	2	O868H5	O868H5 branchiosteo
35	615	13.1	581	2	O9XZM7	O9XZM7 strongyloce
36	609	13.0	688	2	O868H6	O868H6 branchiosteo
37	596	12.7	490	2	O7TND4	O7TND4 mus musculus
38	596	12.7	490	2	O920K3	O920K3 rattus norv
39	595	12.7	490	1	TMS2_MOUSE	O9168 mus musculus
40	594.5	12.7	490	2	O6P7D7	O6P7D7 rattus norv
41	594.5	12.7	490	2	AAH61712	AAH61712 rattus norv
42	594	12.7	688	2	O868H4	O868H4 branchiosteo
43	589.5	12.6	701	2	O9JUS9	O9JUS9 rattus norv
44	583	12.4	422	1	DESI_HUMAN	O9152 homo sapien
45	583	12.4	423	2	O6UW31	O6UW31 homo sapien
46	583	12.4	423	2	AAQ89376	AAQ89376 homo sapi
47	583	12.4	703	2	O8CHN8	O8CHN8 rattus norv
48	582.5	12.4	704	1	CRAR_MOUSE	P88064 mus musculus
49	581.5	12.4	868	2	O9Y1V3	O9Y1V3 polyandroca
50	576.5	12.3	1134	2	O7RTY7	O7RTY7 homo sapien
51	576.5	12.3	1524	2	O91674	O91674 xenopus lae
52	570	12.2	699	1	CRAR_HUMAN	P48740 h complen
53	567	12.1	418	2	O6IEI5	O6IEI5 rattus norv
54	567	12.1	676	2	O6DUU6	O6DUU6 cyprius ca
55	556	11.9	492	2	O96T73	O96T73 homo sapien
56	555.5	11.9	492	2	TMS2_HUMAN	O15393 homo sapien
57	555.5	11.9	492	2	AAH51839	AAH51839 homo sapi
58	554	11.8	681	2	O7ZT70	O7ZT70 lampetra ja
59	549	11.7	698	2	O6GPF9	O6GPF9 xenopus lae
60	548.5	11.7	60	2	O8BM10	O8BM10 mus musculus
61	541	11.6	307	2	O6ZND6	O6ZND6 homo sapien
62	541	11.6	307	2	BAD18439	BAD18439 homo sapi
63	540	11.5	453	2	O812N6	O812N6 mus musculus
64	539	11.5	453	1	TMS3_MOUSE	O8K10 mus musculus
65	538.5	11.5	417	2	O8VH04	O8VH04 rattus norv
66	536.5	11.5	453	2	O6ZWC3	O6ZWC3 homo sapien
67	536.5	11.5	453	2	AAQ88823	AAQ88823 homo sapi
68	536.5	11.5	453	2	BAD18806	BAD18806 homo sapi
69	534	11.4	707	2	O8CGV0	O8CGV0 cyprius ca
70	532.5	11.4	279	2	O9QZ74	O9QZ74 rattus norv
71	532.5	11.4	638	2	O8R0P5	O8R0P5 mus musculus
72	528.5	11.3	618	1	KAL_MOUSE	P6262 mus musculus
73	527.5	11.3	617	2	O8J1S1	O8J1S1 triaklis scy
74	526	11.2	454	1	TMS3_HUMAN	P57727 homo sapien
75	522.5	11.2	417	2	O8VDV1	O8VDV1 mus musculus
76	522.5	11.2	417	2	O8VHK6	O8VHK6 mus musculus
77	520.5	11.1	279	2	O7TNR3	O7TNR3 mus musculus
78	518.5	11.1	600	2	O7ZTR2	O7ZTR2 xenopus lae
79	515	11.0	603	1	CFAI_MOUSE	O61129 mus musculus
80	515	11.0	686	2	O6LOI9	O6LOI9 gallus galli
81	515	11.0	686	2	AA573179	AA573179 gallus ga
82	514.5	11.0	698	2	O9PU71	O9PU71 xenopus lae
83	514	11.0	604	1	CFAI_RAT	O9UWU3 rattus norv
84	510.5	10.9	697	2	O8CG43	O8CG43 rattus norv
85	510.5	10.9	733	2	O8CD27	O8CD27 mus musculus
86	510.5	10.9	733	2	O920S0	O920S0 mus musculus
87	508.5	10.9	418	1	HATT_HUMAN	O60235 homo sapien
88	508	10.8	455	1	TMS5_MOUSE	O9ER04 mus musculus
89	507.5	10.8	730	1	PLMN_PIG	P14272 rattus norv
90	506.5	10.8	645	2	O6ZWK6	O6ZWK6 sus scrofa
91	504.5	10.8	438	2	BAC85495	BAC85495 homo sapi
92	504.5	10.8	438	2	O8BYR2	O8BYR2 homo sapien
93	504.5	10.8	537	2	O9BVE1	O9BVE1 homo sapien
94	504.5	10.8	581	2	O8CFE0	O8CFE0 mus musculus
95	504	10.8	471	2	O03711	O03711 xenopus lae
96	504	10.8	645	2	O7PWE4	O7PWE4 anopheles g
97	504	10.8	728	2	O96R84	O96R84 homo sapien
98	503.5	10.8	445	2	O8CJ17	O8CJ17 rattus norv
99	503.5	10.8	688	2	O9PVY4	O9PVY4 xenopus lae
100	502.5	10.7				

ALIGNMENTS

RESULT 1
ST14 HUMAN STANDARD; PRT: 855 AA.
ID Q9V5Y6; Q9BS01; Q9H3S0; Q9HB36; Q9HCA3;
AC 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE Suppressor of tumorigenicity 14 (EC 3.4.21.-) (Matrilipase) (Membrane-
type serine protease 1) (MT-SPL) (Proteinase) (Serine protease TADG-15)
DE (Tumor associated differentially-expressed gene-15 protein).
GN Name=ST14; Synonyms=PRSS14, SNCL19, TADG15;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99303581; Pubmed=10373424;
RA Lin C.Y., Anders J., Johnson M., Sang Q.A., Dickson R.B.;
RT "Molecular cloning of cDNA for matrilipase, a matrix-degrading serine
RT protease with trypsin-like activity";
RL J. Biol. Chem. 274:18231-18236(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99432178; Pubmed=10500122;
RA Takeuchi T., Shuman M.A., Craik C.S.;
RT "Reverse biochemistry: use of macromolecular protease inhibitors to
RT dissect complex biological processes and identify a membrane-type
RT serine protease in epithelial cancer and normal tissue";
RL Proc. Natl. Acad. Sci. U.S.A. 96:11054-11061(1999).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Prostate;
RA Yamaguchi N., Mitsui S.;
RT "Molecular cloning of a novel transmembrane serine protease expressed
RT in human prostate";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Tanimoto H., Underwood L.J., Wang Y., Shigemasa K., Parnley T.H.,
RA O'Brien T.J.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC TISSUE=Blood, and Muscle;
RX MEDLINE=22388257; Pubmed=12477932; DOI=10.1073/pnas.242603899;
RA Krausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marins K., Farmer A.A., Rubin G.M., Hong L.,
RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,
RA Kana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Pehey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Buterfield J.S.N., Krzywinski M.I., Skalek U., Small D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Maria M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [6]
RP SEQUENCE OF 340-664 FROM N.A.
RA Cao J., Fan W., Zheng S.;
RT "Genomic analysis of a novel human serine protease SMC19";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [7]

RP CHARACTERIZATION.
RC TISSUE=Milk;
RX MEDLINE=99303582; Pubmed=10373425;
RA Lin C.Y., Anders J., Johnson M., Dickson R.B.;
RT "Purification and characterization of a complex containing matrilipase
RT and a Kunitz-type serine protease inhibitor from human milk";
RL J. Biol. Chem. 274:18237-18242(1999).
CC -1- FUNCTION: Degrades extracellular matrix. Proposed to play a role
CC in breast cancer invasion and metastasis. Exhibits trypsin-like
CC activity as defined by cleavage of synthetic substrates with Arg
CC or lys as the P1 site.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
CC -1- SIMILARITY: Belongs to peptidase family S1.
CC -1- SIMILARITY: Contains 2 CUB domains.
CC -1- SIMILARITY: Contains 4 LDL-receptor class A domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF118224; AAD42765.2; -;
DR EMBL; AF133086; AAF00109.1; -;
DR EMBL; AB030036; BAB20376.1; -;
DR EMBL; AF057145; AAG15395.1; -;
DR EMBL; AF005826; AAH05826.2; -;
DR EMBL; BC030532; AAH30532.1; -;
DR EMBL; AF283256; AAG13949.1; -;
DR PDB; 1EAW; X-ray; A/C=615-855.
DR MEROPS; S01.302; -;
DR Genew; HGNC:11344; ST14.
DR MIM; 606797; -;
DR GO; GO:0005887; C:integral to plasma membrane; TAS.
DR GO; GO:0008236; F:serine-type peptidase activity; TAS.
DR GO; GO:0006508; P:proteolysis and peptidolysis; TAS.
DR InterPro; IPR000859; CUB.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR InterPro; IPR009003; Pept_Ser_Cys.
DR Pfam; PF00431; CUB; 2.
DR Pfam; PF00057; Ldl_recept_a; 4.
DR Pfam; PF00089; Trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00261; LDLRCPTOR.
DR SMART; SM00042; CUB; 2.
DR SMART; SM00192; LDla; 4.
DR SMART; SM00020; TRYP_SPC; 1.
DR PROSITE; PS01180; CUB; 2.
DR PROSITE; PS01209; LDla_1; 2.
DR PROSITE; PSS0068; LDla_2; 4.
DR PROSITE; PSS0240; TRYPsin_DOM; 1.
DR PROSITE; PSS0134; TRYPsin_HIS; 1.
DR PROSITE; PS00135; TRYPsin_SER; 1.
KW 3D-structure; Glycoprotein; Hydrolase; Repeat; Serine protease;
KW Signal-anchor; Transmembrane.
FT DOMAIN 1 55 Cytoplasmic (Potential).
FT TRANSMEM 56 76 Signal-anchor for type II membrane
FT 77 855 protein (Potential).
FT DOMAIN 214 334 Extracellular (Potential).
FT DOMAIN 340 447 CUB 1.
FT DOMAIN 452 487 CUB 2.
FT DOMAIN 487 524 LDL-receptor class A 1.
FT DOMAIN 524 560 LDL-receptor class A 2.
FT DOMAIN 566 603 LDL-receptor class A 3.
FT DOMAIN 615 854 LDL-receptor class A 4.
FT ACT_SITE 656 Serine protease.
FT ACT_SITE 656 Charge relay system (By similarity).
FT ACT_SITE 711 Charge relay system (By similarity).
FT ACT_SITE 805 Charge relay system (By similarity).
FT ACT_SITE 805

FT CARBOHYD 109 109 N-linked (GlcNAc...) (Potential) .
 FT CARBOHYD 302 302 N-linked (GlcNAc...) (Potential) .
 FT CARBOHYD 485 485 N-linked (GlcNAc...) (Potential) .
 FT CARBOHYD 772 772 N-linked (GlcNAc...) (Potential) .
 FT CONFLICT 381 381 R -> S (in Ref. 4) .
 FT CONFLICT 674 674 A -> V (in Ref. 3) .
 SQ SEQUENCE 855 AA; 94769 MW; 26143132C01F99C9 CRC64;

Query Match 99.5%; Score 4659; DB 1; Length 855;
 Best Local Similarity 99.6%; Pred. No. 1e-313;
 Matches 852; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MGSDRARKGGGPGDGFAGLKYNSRHEKYNGLIEGVELEPVNNYKYEKGHPGMVYLA 60
 DB 1 MGSDRARKGGGPGDGFAGLKYNSRHEKYNGLIEGVELEPVNNYKYEKGHPGMVYLA 60
 QY 61 VLIGLLVLIGLIGFLVWHLQYRDVVOVKXNGYRITNENFVDAVENSNTSEFVSLASKV 120
 DB 61 VLIGLLVLIGLIGFLVWHLQYRDVVOVKXNGYRITNENFVDAVENSNTSEFVSLASKV 120
 QY 121 KDALKLYSGVPLGPHKESAVTAFSGSYIATYMSFSIPQHLVEAEKVAEERVYM 180
 DB 121 KDALKLYSGVPLGPHKESAVTAFSGSYIATYMSFSIPQHLVEAEKVAEERVYM 180
 QY 181 LPPRARSLSKSVYTSVVAFTDSKTVORTODNSCSFGLHAGVELEMRPTTGPFDPSPYPA 240
 DB 181 LPPRARSLSKSVYTSVVAFTDSKTVORTODNSCSFGLHAGVELEMRPTTGPFDPSPYPA 240
 QY 241 HARCOMALRGDADSVLSLTFRSPDLASCDEBGSJLVTVYNTLSPEMPEHALVOLGTYPPS 300
 DB 241 HARCOMALRGDADSVLSLTFRSPDLASCDEBGSJLVTVYNTLSPEMPEHALVOLGTYPPS 300
 QY 301 YNLFTHSSONVLLTLITNTRRRHGPATFQLPKRNSSCCGRRLKRAQSTNSYYRGHY 360
 DB 301 YNLFTHSSONVLLTLITNTRRRHGPATFQLPKRNSSCCGRRLKRAQSTNSYYRGHY 360
 QY 361 PNPDICWNINIEVPNNQHVKAFFKFFYLLEPGVPATGCKDVEYENGEKYECSERQFVYTS 420
 DB 361 PNPDICWNINIEVPNNQHVKAFFKFFYLLEPGVPATGCKDVEYENGEKYECSERQFVYTS 420
 QY 421 NSNKITVRPHSDQSYDTGTFLAELYSYSDSPCPQOFTCRTRGCIKRELKCDGNADCTDH 480
 DB 421 NSNKITVRPHSDQSYDTGTFLAELYSYSDSPCPQOFTCRTRGCIKRELKCDGNADCTDH 480
 QY 481 SDELINCSGDAGHOTCKRKKFKPKLPWVCDVNDGDNDSDEGCCCPAOTFPCSNKCLSK 540
 DB 481 SDELINCSGDAGHOTCKRKKFKPKLPWVCDVNDGDNDSDEGCCCPAOTFPCSNKCLSK 540
 QY 541 SQOCSNGDGDGSDGDEASCPRVNVYTCRHTYRCCLNGCLSKNPECDGKEDCSGSDSK 600
 DB 541 SQOCSNGDGDGSDGDEASCPRVNVYTCRHTYRCCLNGCLSKNPECDGKEDCSGSDSK 600
 QY 601 DCDGGLRSFTRQARVVGTDADBEWPMQVSLHALGQGHICGASLISPMVLVSAHAYCID 660
 DB 601 DCDGGLRSFTRQARVVGTDADBEWPMQVSLHALGQGHICGASLISPMVLVSAHAYCID 660
 QY 661 DRGRYSDPTQMTWTFGLIHDSQSASAGVQERRLKRIISHPFNDPTFDYIALLEKXP 720
 DB 661 DRGRYSDPTQMTWTFGLIHDSQSASAGVQERRLKRIISHPFNDPTFDYIALLEKXP 720
 QY 721 AEVSMWRPCLTPASHVFPAGKAIWYGMGHTQYGGGALLKKGELRVNNTTCENTL 780
 DB 721 AEVSMWRPCLTPASHVFPAGKAIWYGMGHTQYGGGALLKKGELRVNNTTCENTL 780
 QY 781 PQOITPRMVCVGLSGVDSCQSDSGGFLSVSEADGRIFGAGVYVWGDGCAGRKKPGYTT 840
 DB 781 PQOITPRMVCVGLSGVDSCQSDSGGFLSVSEADGRIFGAGVYVWGDGCAGRKKPGYTT 840
 QY 841 RLPLFRMIXENTGV 855
 DB 841 RLPLFRMIXENTGV 855

RESULT 2
 ID ST14 MOUSE STANDARD; PRT; 855 AA.
 AC P56677;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Suppressor of tumorigenicity 14 (EC 3.4.21.-) (Epithin).
 GN Name=St14; Synonym=Pres14;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C.B.173CID; TISSUE=Thymus;
 RX MEDLINE=99216440; PubMed=1019918;
 RA Kim M.G., Chen C., Lyu M.S., Cho E.G., Park D., Kozak C.,
 RA Schwartz R.H.;
 RT "Cloning and chromosomal mapping of a gene isolated from thymic
 RT stromal cells encoding a new mouse type II membrane serine protease,
 RT epithin, containing four LDL receptor modules and two CUB domains.";
 RL Immunogenetics 49:420-428(1999).
 RN [2]
 RP REVISIONS TO 23; 321; 325; 343; 409-410 AND C-TERMINUS.
 RC STRAIN=C.B.173CID; TISSUE=Thymus;
 RA Kim M.G., Chen C., Cho E.G., Park D., Schwartz R.H.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Breast tumor;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,
 RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.C.,
 RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Hitting M., Madan A., Young A.C., Shcherenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Gilmwood J., Schmutz J., Myers R.M.,
 RA Butcherfield Y.S.N., Krzywinski M.I., Skalak U., Smallie D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
 CC -1- TISSUE SPECIFICITY: Highly expressed in intestine, kidney, lung,
 CC and thymus. Not expressed in skeletal muscle, liver, heart, testis
 CC and brain.
 CC -1- SIMILARITY: Belongs to peptidase family S1.
 CC -1- SIMILARITY: Contains 2 CUB domains.
 CC -1- SIMILARITY: Contains 4 LDL-receptor class A domains.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL: AF042822; AA02230.3; -;
 DR EMBL: BC005496; AA05496.1; -;
 DR HSSP: P00760; 1EXZ.
 DR MEROPS: S01.302; -;
 DR MGD: MGI:133881; St14.

```
DR GO; GO:0005576; C:extracellular; IDA.
DR GO; GO:0019897; C:extrinsic to plasma membrane; IDA.
DR GO; GO:0008236; F:serine-type peptidase activity; IDA.
DR InterPro: IPR000859; CUB.
DR InterPro: IPR001172; LDL_receptor_A.
DR InterPro: IPR001254; Peptidase_S1.
DR InterPro: IPR001314; Peptidase_S1A.
DR Pfam: PF00431; CUB; 2.
DR Pfam: PF00057; LDL_recept_a; 4.
DR Pfam: PF00089; Trypsin_1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR PRINTS: PR00261; LDLRECEPTOR.
DR SMART: SM00042; CUB; 2.
DR SMART: SM00192; LDLa; 4.
DR SMART: SM00020; TRYP_spc; 1.
DR PROSITE: PS01180; CUB; 2.
DR PROSITE: PS01209; LDLR_1; 2.
DR PROSITE: PS50068; LDLR_2; 4.
DR PROSITE: PS50240; TRYPsin_DOM; 1.
DR PROSITE: PS00134; TRYPsin_HIS; 1.
DR PROSITE: PS00135; TRYPsin_SER; 1.
KW Glycoprotein; Hydrolase; Repeat; Serine protease; Signal-anchor;
Transmembrane.
FT DOMAIN 1 55 Cytoplasmic (Potential).
FT TRANSMEM 56 76 Signal-anchor for type II membrane
protein (Potential).
FT 77 855 Extracellular (Potential).
FT DOMAIN 214 331 CUB 1.
FT DOMAIN 340 444 CUB 2.
FT DOMAIN 451 488 LDL-receptor class A 1.
FT DOMAIN 489 522 LDL-receptor class A 2.
FT DOMAIN 523 561 LDL-receptor class A 3.
FT DOMAIN 565 604 LDL-receptor class A 4.
FT DOMAIN 615 854 Serine protease.
FT ACT_SITE 656 656 Charge relay system (By similarity).
FT ACT_SITE 711 711 Charge relay system (By similarity).
FT ACT_SITE 805 805 Charge relay system (By similarity).
FT CARBOHYD 107 107 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 302 302 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 365 365 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 421 421 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 489 489 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 772 772 N-linked (GlcNAc...) (Potential).
SQ SEQUENCE 855 AA; 94654 MW; 4F10E84DA2146DD5 CRC64;

Query March 82.9%; Score 3884; DB 1; Length 855;
Best Local Similarity 81.5%; Pred. No. 4.6e-260;
Matches 697; Conservative 73; Mismatches 85; Indels 0; Gaps 0;

QY 1 MGSDRARKGGGQPDGAGLKYNSRHKVGLBEGVEFLPVNNVKKYKKGPGRWVLA 60
DB 1 MGSNRKAGKAGGSGDFFGAGLKYNSRLNMGFBEGVEFLPVNNAKYKKGPRWVLA 60
QY 61 VLIGLLVLLGIGLVNHLQRYRVQVKNKGYKRTINEFVDAYENSNGTEFVSLASKV 120
DB 61 VLSFLLLSLMAAGLVNHFYRNVRVQKVFNGHRLTNEIFLDAYENSTSEFISLASQV 120
QY 121 KDLKLLSYGVFLGPHYKESAVTAFSEGSVITAYMESEFSIPOHLYEAEKVVAAEERVM 180
DB 121 KEAKLLIYNEVPVLPGRHKSAVTAFFSEGSVITAYMESEFSIPLPLAEEVDANAAVERVVT 180
QY 181 LPPRARSLSKSFVVTSVVAFPTDSKTVQRTODNSCSFGILHARGVLEMFETTPGPPDSPYPA 240
DB 181 LPPRARSLSKSFVLTSSVAFPTDPMQLQRTODNSCSFPLHGAHVTRFTTPGFPNSPYPA 240
QY 241 HARQWALRGDADSVLSLTFRSPFLASCDEKSGSLVTVNTLSMEPHALVOLCGTYPSP 300
DB 241 HARQWALRGDADSVLSLTFRSPFLVAPCDEKSGSLVTVNTLSMEPHAVVRLCGTSPSP 300
QY 301 YNLTFFSSONVLTLLTNTERRHGPGEATPFQDPRKSSCGGRKAKQGFNSPYRPGHY 360
DB 301 YNLTFFSSONVLTLLTNTDRHRPGFEATPFQDPRKSSCGGFLSDTQGTFFSSPYRPGHY 360
```

```
QY 361 PNIDCTWNIEVBNQHVKRFKFFYLEBGPAGTCPDQYVEINKEKTCGEESQFVVT 420
DB 361 PNINICTWNIKVNNNNVKKRFLFLVDPNVNVSCTKDQYVEINKEKTCGEESQFVVS 420
QY 421 NSKQITVRFHSDSGSYDTDTGLAFYLSYDSDDPCPGQFTCTGTGCIKEKLRCDMACTDH 480
DB 421 NSKQITVRFHSDSGSYDTDTGLAFYLSYDSNDPCPGMCKTGACIKREKRCDMAADCPDY 480
QY 481 SDELINSCDAGHOFTCKNKKCKPLFWVCDSVNDCCGNSDEOGSCCPAQTRFCSNGKLSK 540
DB 481 SDERYRCNATHQFTCKNQCKPLFWVCDSVNDCCGNSDEOGSCCPAGSTRKCSNGKCLPQ 540
QY 541 SQCKNGKNDGSDGSDASCPKVVVVTCTKATYKCLNGCLCKSKNPEDCKEDCSGSDX 600
DB 541 SQCKNGKNDGSDGSDASCVSVVVSCTKYTYKQNGCLCKSKNPEDCKEDCSGSDX 600
QY 601 DCCGGLRSPFRQARVVGCTDADGEMPMQVSLHALCQGHICGASLISPMVLVSAACFYD 660
DB 601 NCDCGLRSPFRQARVVGCTDADGEMPMQVSLHALCQGHICGASLISPMVLVSAACFYD 660
QY 661 DRGFRYSDPTQWTAFLGLHDQSORSAAGVQERLKRILSHFFNDFTFDYDIALLELEKP 720
DB 661 DKAFKISDYTMWTAFLGLHDQSORSAAGVQELKRLIITHPSFNDTFDYDIALLELEKS 720
QY 721 AESSMVRPILCPDASHVPPAGRAIWTGNGHTQYSGTGLILQKEIRVINOCTCENTIL 780
DB 721 VESTVTRPILCPDATHVFPAGRAIWTGNGHTKEGTGLILQKEIRVINOCTCEDILM 780
QY 781 PQGITPRMCKVGLSGVDSQCGSGPLSSVADRRIGAGVSVSGDAGKNGKGYVT 840
DB 781 PQGITPRMCKVGLSGVDSQCGSGPLSSVADRRIGAGVSVSGDAGKNGKGYVT 840
QY 841 RLPLFRDWIKENTGV 855
DB 841 RLPLFRDWIKENTGV 855

RESULT 3
BAC35125 PRELIMINARY; PRT; 855 AA.
AC BAC35125.
DT 14-APR-2004 (TEMBLrel. 27, Created)
DT 14-APR-2004 (TEMBLrel. 27, Last sequence update)
DT 14-APR-2004 (TEMBLrel. 27, Last annotation update)
DE 0 day neonate kidney cDNA, RIKEN full-length enriched library,
DE clone: D630041B06 product: suppression of tumorigenicity 14 (colon
DE carcinoma), full insert sequence.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Kidney;
RX MEDLINE=22354683; PubMed=1246851;
RA The RIKEN Genome Consortium;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Kidney;
RX MEDLINE=21085660; PubMed=11217851;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690 (2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Kidney;
RX MEDLINE=99279253; PubMed=10349636;
RA Carninci P., Hayashizaki Y.;
```

RT "High-efficiency full-length cDNA cloning."
 RL Mech. Enzymol. 303:19-44(1999).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Kidney;
 RX MEDLINE=20499374; PubMed=11042159;
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
 RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to
 RT prepare full-length cDNA libraries for rapid discovery of new genes."
 RL Genome Res. 10:1617-1630(2000).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Kidney;
 RX MEDLINE=20530913; PubMed=11076861;
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
 RA Kono H., Akiyama Y., Nishi K., Katsuna T., Tashiro H., Itoh M.,
 RA Suni N., Ishii Y., Nakamura S., Hazama N., Nishine T., Harada A.,
 RA Yamamoto S., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
 RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watabiki M.,
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format
 RT sequencing pipeline with 384 multicapillary sequencer."
 RL Genome Res. 10:1757-1771(2000).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Kidney;
 RA Adachi J., Aizawa K., Akimura T., Hara A., Hashizume W.,
 RA Fukuda S., Furuno M., Hanagata T., Hara A., Hashizume W.,
 RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
 RA Hori F., Imocani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
 RA Katoh H., Kawai J., Kojima Y., Kondo S., Kono H., Kouda M., Koya S.,
 RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
 RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Sano H.,
 RA Saito R., Saitoh H., Sakai K., Sakazume N., Sano H.,
 RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
 RA Tagawa A., Takahashi F., Takaku-Akashira S., Takeda Y., Tanaka T.,
 RA Tomaru A., Toya T., Yaeunishi A., Muramatsu M., Hayashizaki Y.,
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK052738; BAC35125.1; -
 SQ SEQUENCE 855 AA; 94654 MW; 4F10B84DA2146DD5 CRC64;
 Query March 82.9%; Score 3884; DB 2; Length 855;
 Bees Local Similarity 81.5%; Pred. No. 4.6e-260;
 Matches 697; Conservative 73; Mismatches 85; Indels 0; Gaps 0;
 QY 1 MGSDRARKGGGPDGAGLKYNSRHEKNGLEGEVEPLPVNNVYKVEKPGRVVLA 60
 DB 1 MGSNRGRAGGSGDFFAGLKYNSRLNMGFEVEVELPANNAKVKGKGRPVVLA 60
 QY 61 VLIGLLVLVIGLIGLVHLYQYRDVYKQKQKMTINENFVDAVENSSTPEVSLASKV 120
 DB 61 VLFSLILSLMAAGLLVHFHYRNVAQKVPNGHLRTIBFLDAVENSTSTPEVSLASQV 120
 QY 121 KDAKLKLYSGVPFGCPYHESAVYAFSGSVIAYVWSFSPQHLYEAEVMAEERVYM 180
 DB 121 KEAKLILNEVPVIGPYHKSAVTAFASSGSVIAIYWSFSPHIAEEVDRAAAERVVT 180
 QY 181 LPPARSLKSLFVYTSVVAFPDTSKTVQRTONSCSFGIHAAGVELMTGTFGPPSPYPA 240
 DB 181 LPPARSLKSLFVYTSVVAFPDPRLOKRTONSCSFGIHAAGAVTRTGTGFPSPYPA 240
 QY 241 HARCQWALRGADSVLSTFRSPDLASCDERGSGLVTVYNTLSPMEPHALVOLCGTTPS 300
 DB 241 HARCQWALRGADSVLSTFRSPDLASCDERGSGLVTVYNTLSPMEPHAVRLCGTTPS 300
 QY 301 YNLTFHSSQNVLLTLITNTERBRHGFATFPOLPRMSCGGRLLKAGCTNSPYPGHY 360
 DB 301 YNLTFHSSQNVLLTLITNTERBRHGFATFPOLPRMSCGGFLSDTGTSSPYPGHY 360
 QY 361 PPNDICTNINIEVPNNQHVAFKFFVLLPGVPACTCPKDYVEINGEKYCGERSQFVTS 420

DB 361 PPNDICTNINIKVPNNRAVYKFKLFLYVDENVVSGCTKDYVEINGEKYCGERSQFVSS 420
 QY 421 NSNKTITRPSDOSYPTDTGLAEYLSYSSDPGCGCTCTGRCIRKELBCDGAADCTDH 480
 DB 421 NSSKITVYFHSDBSHYDTGTGLAEYLSYSSNDPCGMFMCKTGRCIRKELBCDGAADCPDY 480
 QY 481 SDEINSCSDAGHOFCTCNKCKPFLFWCDSDVNDGSDSDGSCSPAQTRFCSNGKCLSK 540
 DB 481 SDEHYCCNMTHTQTCNQCCKPLFWCDSDVNDGSDSDGSCSPAGSRKCSNGKCLPQ 540
 QY 541 SQQCNKGDCGSDGSDASCPKRVNVVTCTKHTYRCLNGLCLSKNGPECDGKDCSDGDEK 600
 DB 541 SQKCNKGDCGSDGSDASCSVNVVSCCTKYTRCONGLCLSKNGPECDGKDCSDGDEK 600
 QY 601 DCDGLASFTROARVVGSTADDEGEWQVSLHALGCGHLCGASLSPNMLVSAHICYD 660
 DB 601 NCDGLASFTROARVVGSTADDEGEWQVSLHALGCGHLCGASLSPNMLVSAHICYD 660
 QY 661 DRGFRSDPTQWTFLLGHPQSRAPGVORRLKRIISHPFNDFTFYDIALLELEKP 720
 DB 661 DKMKYSDYTWKTFLLGLDQSRKASGVDELKRIITHPSPNDFTFYDIALLELEKS 720
 QY 721 AEVSMVRPCLCPDASHVPAGKAIWVTGWHTQYGGTGAALLQKGEIRVINQTCENLL 780
 DB 721 VEYSTVVRPCLCPDATHVPAGKAIWVTGWHTGEGTGAALLQKGEIRVINQTCEDLM 780
 QY 781 PQQITPRMCCVGLSGVDSCQDGGPGLSSVADGRIFGAGVYMWKDGCAGRKPGVYT 840
 DB 781 PQQITPRMCCVGLSGVDSCQDGGPGLSSVADGRIFGAGVYMWKDGCAGRKPGVYT 840
 QY 841 RLPLFRDMIKENTGV 855
 DB 841 RLPLFRDMIKENTGV 855
 RESULT 4
 Q9JUI7 PRELIMINARY; PRT; 855 AA.
 ID Q9JUI7
 AC Q9JUI7
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE Membrane bound serine protease (Membrane bound arginine specific
 DE serine protease).
 GN Name=MBSP.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=wistar; TISSUE=Jejunum;
 RX MEDLINE=21456307; PubMed=11573963;
 RA Sato M., Yamaguchi Y., Tanizaki S., Hitomi Y., Iwanaga T., Fushiki T.;
 RT "A role for membrane-type serine protease (MT-Sp1) in intestinal
 RT epithelial turnover."
 RL Biochem. Biophys. Res. Commun. 287:995-1002(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=wistar; TISSUE=Duodenum;
 RX Inoue H., Takahashi K., Kishi K.;
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 CC -1; SIMILARITY: Belongs to peptidase family S1.
 DR EMBL; AB037898; BAB03502.1; -
 DR EMBL; AB049189; BAB13765.1; -
 DR PIR; JCT731; JCT731.
 DR HSSP; P00760; IEZX.
 DR MEROPS; S01.302; -
 DR GO; GO:0004263; F:chymotrypsin activity; IEA.
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0004295; F:trypsin activity; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000859; CUB.

DR InterPro: IPR002172; LDL_receptor_A.
 DR InterPro: IPR001254; Peptidase_S1.
 DR InterPro: IPR001314; Peptidase_S1A.
 DR InterPro: IPR009003; Pept_Ser_Cys.
 DR Pfam: PF00431; CUB; 2.
 DR Pfam: PF00057; Ldl_recept_a; 4.
 DR Pfam: PF00089; Trypsin; 1.
 DR PRINTS: PR00722; CHYMOTRYPSIN.
 DR PRINTS: PR00261; LDLRECEPTOR.
 DR PROSITE: PS01180; CUB; 2.
 DR PROSITE: PS01209; LDLRA_1; 2.
 DR PROSITE: PS00068; LDLRA_2; 4.
 DR PROSITE: PS00240; TRYPSIN_DOM; 1.
 DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.
 DR PROSITE: PS00135; TRYPSIN_SER; 1.
 KM Hydrolyase; Protease; Serine protease.
 QO SEQUENCE 855 AA; 94955 MW; 35806B7E6CF6CF03D CRC64;

Query Match 82.6%; Score 3866; DB 2; Length 855;
 Best Local Similarity 80.8%; Pred. No. 8e-259;
 Matches 691; Conservative 79; Mismatches 85; Indels 0; Gaps 0;

QY 1 MGSDBAKGGGGPRDPAGLKYNRHREKNGLEGVFLPVNNKTKYKXGPRGVVLA 60
 DB 1 MGNRGRKAGGSGSDPAGLKYNSRLNMGFEKGVFLPVNNAKQYKXGPRGVVMA 60
 QY VLGILGLVLLGIGFLVWHLQYRDVVOKYKNGYRINTENFVDAEYNSNSTEFLASKV 120
 DB VVPSFLLSLMAGLLVWHFHYRNRIRIQKVFGRHRIINENFLDAKENSTSEFSLASQV 120
 QY 121 KDALKLYSGVPLGPRYHKESAVTAPEEGSVIAYYGEPSIPOHLVEBAERVAEEVVM 180
 DB 121 KEAKLMYSEVPVIGPRYHKESVTAPEEGSVIAYYGEPSIPHLVEEVRAVMAVERVVT 180
 QY 181 LPPRABSLKSVYVSVVAFPTDSKTQVORTDONSFCGLHARGVLMFTTGPSPSYPA 240
 DB 181 LPPRABSLKSVYVSVVAFPTDSKTQVORTDONSFCGLHARGVLMFTTGPSPSYPA 240
 QY 241 HARCQMLRGDAVSLSLTFERSPDASCDEGSDLVVYNTLSPEBPHALVOLCGTTPS 300
 DB 241 HARCQWLKRGDABVLSLTFERSPDVACDGHDSLVVYNTLSPEBPHAVRLOCTGPS 300
 QY 301 YNLTFFSSQVNLITLITNTERRHGEATFPOLPRMSSCGGLRKAQGFNPSYYPGHY 360
 DB 301 YNLTFFSSQVNLITLITNTERRHGEATFPOLPRMSSCGGLRKAQGFNPSYYPGHY 360
 QY 361 PPNIDCTWNIENVNNOVKVRFKFFVLEEGVAPGCPKQVYVLENGKXGSEBQFVYTS 420
 DB 361 PPNIDCTWNIENVNNOVKVRFKFFVLEEGVAPGCPKQVYVLENGKXGSEBQFVYTS 420
 QY 421 NSNKTIVRFHSDSYTDGFLAEVLSYDSDPCPGQFCTGRCIRKELRCGWDCTDH 480
 DB 421 NSNKTIVRFHSDSYTDGFLAEVLSYDSDPCPGQFCTGRCIRKELRCGWDCTDH 480
 QY 481 SDELNCGCDAGHOTCKKKCKKPLFWYCDVYNDGDSDEQSCCPAQTRCSNGKLSK 540
 DB 481 SDELNCGCDAGHOTCKKKCKKPLFWYCDVYNDGDSDEQSCCPAQTRCSNGKLSK 540
 QY 541 SDOENHCRCAVTHQWCKQKQFCKPLFWYCDVYNDGDSDEQSCCPAQTRCSNGKLSK 540
 DB 541 SDOENHCRCAVTHQWCKQKQFCKPLFWYCDVYNDGDSDEQSCCPAQTRCSNGKLSK 540
 QY 541 SDOENHCRCAVTHQWCKQKQFCKPLFWYCDVYNDGDSDEQSCCPAQTRCSNGKLSK 540
 DB 541 SDOENHCRCAVTHQWCKQKQFCKPLFWYCDVYNDGDSDEQSCCPAQTRCSNGKLSK 540
 QY 601 DCDGGLSFTROARVVGTDADGEMPMOYSLHAGCHLCSGLISPNMLVSAHACYID 660
 DB 601 DCDGGLSFTROARVVGTDADGEMPMOYSLHAGCHLCSGLISPNMLVSAHACYID 660
 QY 661 DRGRYSDPQWTAFLGLHDSQSAAGVQERRIKRIISHPFNDFFDYDIALLEKRP 720
 DB 661 DRGRYSDPQWTAFLGLHDSQSAAGVQERRIKRIISHPFNDFFDYDIALLEKRP 720
 QY 721 AEVSSMWRLPCLPASHVPRAGKAIWYTGNGHIOYGGTGLIIQXGIRVYNQTTCNL 780
 DB 721 AEVSSMWRLPCLPASHVPRAGKAIWYTGNGHIOYGGTGLIIQXGIRVYNQTTCNL 780

QY 781 PQQITPRMVCVGLSGGVDSQCGDSCGPLSSVEADGRIFGAVGVSWGDCGAGNRKGVYT 840
 DB 781 PQQITPRMVCVGLSGGVDSQCGDSCGPLSSVEADGRIFGAVGVSWGDCGAGNRKGVYT 840

QY 841 RLPLFRDMIKENTGV 855
 DB 841 RIPEVRDMIKENTGV 855

RESULT 5
 ID Q6GR54 PRELIMINARY; PRT; 845 AA.
 AC Q6GR54;
 DT 05-JUL-2004 (TEMBLrel. 27, Created)
 DT 05-JUL-2004 (TEMBLrel. 27, last sequence update)
 DT 05-JUL-2004 (TEMBLrel. 27, last annotation update)
 DE Scl4-A-prov protein.
 GN Name=scl4-A-prov;
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;
 OC Xenoportidae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Pelngold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Maravita K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loggiano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Guiraud P.H.,
 RA Richardson S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W., Butterfield Y.S.,
 RA Krzyanski M.I., Skalska U., Smalins D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Maxam M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RX MEDLINE=22341132; PubMed=12454917;
 RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
 RA Richardson P.;
 RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
 RT Initiative.";
 RL Dev. Dyn. 225:384-391 (2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RA Klein S., Strausberg R.;
 RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
 CC - SIMILARITY: Belongs to peptidase family S1.
 DR EMBL: BC071077; AAH71077.1; -;
 DR InterPro: IPR000859; CUB.
 DR InterPro: IPR002172; LDL_receptor_A.
 DR InterPro: IPR001254; Peptidase_S1.
 DR InterPro: IPR001314; Peptidase_S1A.
 DR InterPro: IPR009003; Pept_Ser_Cys.
 DR Pfam: PF00431; CUB; 2.
 DR Pfam: PF00057; Ldl_recept_a; 4.
 DR Pfam: PF00089; Trypsin; 1.
 DR PRINTS: PR00722; CHYMOTRYPSIN.

DR PRINTS; PRO0261; LDRRECEPTOR.
 DR SMART; SM00042; CUB; 2.
 DR SMART; SM00192; LDLa; 4.
 DR SMART; SM00020; TRY_SPC; 1.
 DR PROSITE; PS01180; CUB; 2.
 DR PROSITE; PS01209; LDLa_1; 2.
 DR PROSITE; PS50068; LDLa_2; 4.
 DR PROSITE; PS50240; TRYPsin_DOM; 1.
 DR PROSITE; PS00134; TRYPsin_HIS; UNKNOWN_1.
 DR PROSITE; PS00135; TRYPsin_SER; 1.
 DR HydroLase; Protease; Serine protease.
 SQ SEQUENCE 845 AA; 93593 MW; B917386C95BC73BD CRC64;

Query Match 57.4%; Score 2686; DB 2; Length 845;
 Best Local Similarity 55.7%; Pred. No.3.7e-177; Indels 10; Gaps 5;

Matches 472; Conservative 154; Mismatches 212; Indels 10; Gaps 5;

QY 14 KDFGAGLKYSRHKVNGLEGEVEFLPVNNVKKYKHKPGRWVLAVALIGLLVLLGIG 73
 DB 2 KDSQSMKKYNNRQSLNGFEEVEFLPATNSKYKTKPKKKLAIIGLVIGALLSLTIG 61
 QY 74 FLVHLLQYRDVVRQVKNGKRTINENFVDAIENSSTEFVSLASKYKDALIKLYSGVPF 133
 DB 62 LLVWHFAYRNKPVNKKLYTGTYLTIANTPIDAYENSTTAEFSDLSAKVIDTLQTYVNGNKD 121
 QY 134 LGPVHKSATVAFSEG---SVIAYVSEFSI PQHLVEBAERVMAEERVVMLEPRARSLKS 190
 DB 122 IAPLQKCSISAFSEGGNNVIGYVSEFVPAREAAFEAKISLKLPSVNPOR--T 178
 QY 191 FVTVSVVAFPTDSKTVOVRTQDNCSFGIHAARVEMFTTPGFPSFPYAHARCOMALRG 250
 DB 179 FALDSIAVAPTPDPAIVAFKSSCAYFLHSSNGVAVKSFSSGFPDSYPRARAKLMTLRA 238
 QY 251 DADSVLSLTFERSPLASCDEKSGDLVTVYNTLSMEPHALVOLCGTTPSPSYNTLTFHSSQN 310
 DB 239 DAGRIHLHFEKTFMEKCKPMGSGFVWVYDLSLIEPRAQIRLIGIYPPSYNTLTFSSSN 298
 QY 311 VLLITLTNTERRRHGEFATFQFLPRMSSCGGRRLKQGTNSPYRPHYENPINDCMNI 370
 DB 299 VMLVTLYTDVNGKPGFLAEFKQLPKTSLCCGLRDASGRTSPFPAPHYPTSTECIMDI 358
 QY 371 EVPNNAQVKKVAFKFFYLLPEGPVAGTCKPDYVEINGEKYCGERSQFVNTSNNKITVRFH 430
 DB 359 QVPENKFKVKKAFNNFYLAEPGVPTKCTKDVEIKGQYCGEKKFPVSNNSKMSRFFV 418
 QY 431 SDQSYTDTGFLAELYSIDSDPCPGQFTCRGCIKRELKCDGAADCTDHSDELNCSDA 490
 DB 419 SDQSYTDTGFLAELYSIEPRNPCEPDQFCRSGRCIRLDQKCDGNWDCEDFSDENSCCTA 478
 QY 491 GHQFTCKK-KFCKPLFWVCSVNDGNSDQSGSCPAQTRFCNGKCLSKSQCNKDD 549
 DB 479 -LQRCVNSKLCCKSFYPCDGVNDCGSSDLACKCPNNYTKCNGKCIPTSDQCDARDN 537
 QY 550 CGDSDSDASCPKVVAVVCTKATYRCLNGLCISKNPEDCGEKEDSDSDSEK--CCDGLR 607
 DB 538 CGDSDSDASCDQVLTACTEYTKCKNQCITKKNPEDEGSDSDSDSDDEMAACNCGKR 597
 QY 608 SFTROARVVGSTDADEGEWPMQVSLHALVQGHICGASILISNNMLVSAACIYIDRGFRYS 667
 DB 598 PFTKKSRIYGVGNADTEGFPMQVSLHALKGNHGTGASIVSPMLISAACHQDODHQRYS 657
 QY 668 DPQOWTFLGLHDSORSARVQERRLKRITISHPFNDFTEPDVIALLELEKPAEYSMV 727
 DB 658 DASLMTIYLGLHDDAQLNTKDQVVERKIKRIIAHIGFNDNTYDNDIAVLELEKPEVYDFPI 717
 QY 728 RPICLPASHVFPAGKAIWVTGMGHTQYGTGALILQKGEIRVINGTTCENTLIPQOITPR 787
 DB 718 QPVCIPSTHDFPGKEPIWVTGMALKKEGGAIVILQGAELIRIINQRECNLLDQGLTPR 777
 QY 788 MCVGFLSAGVSDCGSDSGPSSVEADGRIFAGGVVSWGDCGGRNKGYYTTLPLFRD 847
 DB 778 MLCAGVSGGIDACCGDSDSGPSSVEALNNKYLAGVNVSWGEGCARRNKPGYYTIVSNMRD 837

QY 848 WIKENTGV 855
 DB 838 WIKDXTGL 845

RESULT 6
 ID Q9DGR1 PRELIMINARY; PRT; 845 AA.

AC Q9DGR1; 01-MAR-2001 (TREMblrel. 16, Created)
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
 DT 01-MAR-2004 (TREMblrel. 26, Last annotation update)
 DE Homolog of human MT-SPI.

GN Name:XMt-SPI;
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
 OC Xenopodinae; Xenopus.
 NX NCBI_Taxid=8355;
 RN [1]

RP SEQUENCE FROM N.A.
 RA MEDLINE=20363741; PubMed=10903452;
 RX Yamada K., Takabatake T., Takeshima K.;

RT "Isolation and characterization of three novel serine protease genes
 from Xenopus laevis."

RL Gene 252:209-216 (2000).
 CC -1 SIMILARITY: Belong to peptidase family S1.

DR EMBL; AB038498; BAB08218.1; -.
 DR HSSP; P00760; 3BTH.

DR MEROPS; S01.050; -.
 DR GO; GO:0004263; F:chymotrypsin activity; IEA.

DR GO; GO:0004295; F:trypsin activity; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.

DR InterPro; IPR002172; LDLa_receptor_A.
 DR InterPro; IPR001254; Peptidase S1.

DR InterPro; IPR001314; Peptidase S1A.
 DR InterPro; IPR009003; pept_Ser_Cys.

DR Pfam; PF00431; LDLa_2.
 DR Pfam; PF00057; LDLa_recept_a; 4.

DR Pfam; PF00089; Trypsin_1.
 DR PRINTS; PRO00722; CHYMOTRYPSIN.

DR PRINTS; PRO0261; LDRRECEPTOR.
 DR SMART; SM00042; CUB; 2.

DR SMART; SM00192; LDLa; 4.
 DR SMART; SM00020; TRY_SPC; 1.

DR PROSITE; PS01180; CUB; 2.
 DR PROSITE; PS01209; LDLa_1; 2.

DR PROSITE; PS50068; LDLa_2; 4.
 DR PROSITE; PS50240; TRYPsin_DOM; 1.

DR PROSITE; PS00134; TRYPsin_HIS; UNKNOWN_1.
 DR PROSITE; PS00135; TRYPsin_SER; 1.

KW HydroLase; Protease; Serine protease.
 SQ SEQUENCE 845 AA; 93597 MW; 7FD7E62851A758B7 CRC64;

Query Match 56.8%; Score 2660; DB 2; Length 845;
 Best Local Similarity 55.4%; Pred. No.2.4e-175; Indels 10; Gaps 5;

Matches 470; Conservative 155; Mismatches 216; Indels 10; Gaps 5;

QY 14 KDFGAGLKYSRHKVNGLEGEVEFLPVNNVKKYKHKPGRWVLAVALIGLLVLLGIG 73
 DB 2 KDSQSMKKYNNRQSLNGFEEVEFLPATNSKYKTKPKKKLAIIGLVIGALLSLTIG 61
 QY 74 FLVHLLQYRDVVRQVKNGKRTINENFVDAIENSSTEFVSLASKYKDALIKLYSGVPF 133
 DB 62 LLVWHFAYRNKPVNKKLYTGTYLTIANTPIDAYENSTTAEFSDLSAKVIDTLQTYVNGNKD 121
 QY 134 LGPVHKSATVAFSEG---SVIAYVSEFSI PQHLVEBAERVMAEERVVMLEPRARSLKS 190
 DB 122 IAPLQKCSISAFSEGGNNVIGYVSEFVPAREAAFEAKISLKLPSVNPOR--T 178
 QY 191 FVTVSVVAFPTDSKTVOVRTQDNCSFGIHAARVEMFTTPGFPSFPYAHARCOMALRG 250

Db 179 FALDSLVAEPTDPOIAFVFNKNSCAVFLHSSNGVAFSSPSPGPDSPYPNARCLMTLIRA 238
 Qy 251 DADSVSLFTRSPFLASCDERGSGLVYVNTLSMPHEHALVOLCGTYPSPYNTLTHSSON 310
 Db 239 DAGRIHLHRTKFMCKKPNGBFVNVYDLSLTERPAQRLGILPSPSTLTFSSSN 298
 Qy 311 VLLITLITNERRRHPGEATFPOLRPMSSCGRLRKAQGTNSPYPYGHVPPNIDCTWNI 370
 Db 299 VMLVTLVTDVNGKPGFLAEKQLPKTSLGGILRDASGFTSPYPAHYPPSPSTESIMDI 358
 Qy 371 EVPRNNGVAKRPFKFFYLLEGGVAGTGPXYVEINSGKTCGERSQFVYTSNKKITTFRH 430
 Db 359 QVBDNKKFVKRPFMVFYAEKGVVPTKCTDVEIKGQYGBEKFVVSNNSSKMSYRFV 418
 Qy 431 SDQSYDTGFLAEFLVSDSDPCPGQGTCTGRCIRKELCDDGADCTDSDDELNCSDA 490
 Db 419 SDQSYDTGFLAEFLVSDSDPCPGQGTCTGRCIRKELCDDGADCTDSDDELNCSCTA 478
 Qy 491 GHOFTCKN-KECKPLFWVCDSDVNDGDNSSDEGSCSPAQTFRSGNCKLSKQOQCNKGD 549
 Db 479 -LQPRCVNSHLCKRSPYICDGVNDGSDSDELAACKCPNNPFKCGNGKCIDPSQKCDRVND 537
 Qy 550 CGRGSDBASCPKNNVVTCTHTYRCLNGLCLSKNPECDKEDCSGSDSK--DCDQGLR 607
 Db 538 CGRGSDBASCPKNNVVTCTHTYRCLNGLCLSKNPECDKEDCSGSDSK--DCDQGLR 597
 Qy 608 SPTROARVVGCTDADGEWMPQVSLHALGQGHICGASLSPNMLVSAHVCYIDDRGFRYS 667
 Db 598 PFTKKSIVGVANADTEPFMNOVSLHAKNKHCTGASLGPFTMLISAAHCFQDDHQKRS 657
 Qy 668 DPTQWTAFLIGLHDSQSRABGVDERLKRILSHPFNDFEDYDIALLEKPAEYSSMV 727
 Db 658 DASIMTAYLGLHDOAQNTKDVERRIKRIMAHIGFNDNTYDNDIAVLEKPEVETDPI 717
 Qy 728 RPTCLPASHVFPAGKAIWNTGKHTGYGSGALILQKGEIRVINTOTTENLPLPQUTPR 787
 Db 718 QPVCIPSTHDFPVGKPIWGTGALKEGGAVALILQKGEIRVINTOTTENLPLPQUTPR 777
 Qy 788 MMCVGFSLSGVDSQCGSDGPGPLSSVEADGRIFGAVVSWMGDCAGRRKPGVYTLPLFRD 847
 Db 778 MLCAGFVSGSIDACQSDSGPPLSSVELNNKYVLAGVSWEGCARRKPGVYTLVSMRD 837
 Qy 848 WIKENTGV 855
 Db 838 WSKDKTGL 845
 RESULT 7
 Q8WVC1 PRELIMINARY; PRT; 422 AA.
 AC Q8WVC1; 08WVC1; 20, Created)
 DT 01-MAR-2002 (Tremblrel, 20, Last sequence update)
 DT 01-MAR-2004 (Tremblrel, 26, Last annotation update)
 DE ST14 protein (Fragment).
 GN Name=ST14;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NC NCB1; TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Colon;
 RX MEDLINE=2328257; PubMed=12477932;
 RA Straussberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marsina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udén T.B., Toshitsuki S., Carninci P., Prange C.,
 RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huijck S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren B.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Ketterman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W., Butterfield Y.S.,
 RA Krzyzanski M.I., Skalska U., Smalins D.E., Scherch A., Schein J.E.,
 RA Jones S.J., Maria M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Colon;
 RA Straussberg R.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: Belongs to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC018146; AAH18146.1; -;
 DR GO; GO:0004263; F:chymotrypsin activity; IEA.
 DR GO; GO:0004295; F:trypsin activity; IEA.
 DR GO; GO:0005508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR002172; LDL_receptor_A.
 DR InterPro; IPR001254; Peptidase_S1.
 DR InterPro; IPR001314; Peptidase_S1A.
 DR InterPro; IPR009003; Pept_Ser_Cys.
 DR Pfam; PF00057; Ldl_recept_a; 4.
 DR Pfam; PF00089; Trypsin_1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR PRINTS; PR00261; LDLRECEPTOR.
 DR SMART; SM00192; LDLa; 4.
 DR SMART; SM00020; TYP_SPC; 1.
 DR PROSITE; PS01209; LDLRA_1; 2.
 DR PROSITE; PS00068; LDLRA_2; 4.
 DR PROSITE; PS50240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR Hydrolase; Protease; Serine protease.
 KW NON TER
 FT
 SQ SEQUENCE 422 AA; 46257 MW; 2C99875D1B588319 CRC64;
 Query Match 50.5%; Score 2365; DB 2; Length 422;
 Best Local Similarity 99.5%; Pred. No. 2, 6e-155;
 Matches 420; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Qy 434 SYTDGFLAEFLVSDSDPCPGQGTCTGRCIRKELCDDGADCTDSDDELNCSDAHQ 493
 Db 1 SYTDGFLAEFLVSDSDPCPGQGTCTGRCIRKELCDDGADCTDSDDELNCSDAHQ 60
 Qy 494 FTCKNFKCPLEFVNCSDVNDGDNSSDEGSCSPAQTFRSGNCKLSKQOQCNKGDG 553
 Db 61 FTCKNFKCPLEFVNCSDVNDGDNSSDEGSCSPAQTFRSGNCKLSKQOQCNKGDG 120
 Qy 554 SDEASCPKNNVVTCTHTYRCLNGLCLSKNPECDKEDCSGSDSK--DCDQGLR 613
 Db 121 SDEASCPKNNVVTCTHTYRCLNGLCLSKNPECDKEDCSGSDSK--DCDQGLR 180
 Qy 614 RVVGGTGDADGEWMPQVSLHALGQGHICGASLSPNMLVSAHVCYIDDRGFRYSDEPTQW 673
 Db 181 RVVGGTGDADGEWMPQVSLHALGQGHICGASLSPNMLVSAHVCYIDDRGFRYSDEPTQW 240
 Qy 674 AFLGLHDQSORSPAVGOERRLKRILSHPFNDFEDYDIALLEKPAEYSSWVRDCLP 733
 Db 241 AFLGLHDQSORSPAVGOERRLKRILSHPFNDFEDYDIALLEKPAEYSSWVRDCLP 300
 Qy 734 DASHVPPAGKAIWNTGKHTGYGSGALILQKGEIRVINTOTTENLPLPQUTPRMVCVF 793
 Db 301 DASHVPPAGKAIWNTGKHTGYGSGALILQKGEIRVINTOTTENLPLPQUTPRMVCVF 360
 Qy 794 LSGGVNSCGSDSGPPLSSVEADGRIFGAVVSWMGDCAGRRKPGVYTLPLFRDWTIKENT 853
 Db 361 LSGGVNSCGSDSGPPLSSVEADGRIFGAVVSWMGDCAGRRKPGVYTLPLFRDWTIKENT 420
 Qy 854 GV 855

Db 421 CV 422

RESULT 8

Q6DEVO PRELIMINARY; PRT; 663 AA.

AC Q6DEVO; 01-OCT-2004 (TREMblrel. 28, Created)

DT 01-OCT-2004 (TREMblrel. 28, Last sequence update)

DT 01-OCT-2004 (TREMblrel. 28, Last annotation update)

DE Hypothetical protein.

OS Xenopus tropicalis (Western clawed frog) (Silurana tropicalis).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;

OC Xenopodinae; Xenopus.

OX NCBI_TaxID=8364;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Embryo;

RX MEDLINE=22388257; PubMed=12477932;

RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heieh F., Ditschenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Uedlin T.B., Toshiyuki S., Carrinci P., Prange C., Raha S.S., Loguelli N.A., Peters G.J., Abramson R.D., Mullaly S.J., Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A., Whiteley M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez R.M., Touchman J.W., Schmutz J., Myers R.M., Butterfield Y.S., Krzywnski M.I., Skalska U., Smallos D.E., Scherch A., Schein J.E., Jones S.J., Maira M.A., "Generation and initial analysis of more than 15,000 full-length human RT and mouse cDNA sequences."

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE=Embryo;

RA Klein S., Strauberg R.; Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.

RL EMBL; BC076994; AAH76994.1; -

KM Hypothetical protein.

SQ SEQUENCE 663 AA; 73914 MW; 46B2A56C657C7739 CRC64;

Query Match 42.0%; Score 1965; DB 2; Length 663;

Best Local Similarity 42.9%; Pred. No. 2,2e-127;

Matches 364; Conservative 120; Mismatches 172; Indels 192; Gaps 6;

QY 14 KDPGAGLKYNGRHRKVNLEEGVEFLPVNNKYKYEKGKPGWVLAANTLIGLLVILGIG 73

DB 2 KDSQSMKYNRPQSGMGFEVEGLPIYANTKYEKAPKCKKLIFGVVIAALLSLTIG 61

QY 74 FLVWHLQYRDVQVKNGYRITNTENVDAVENSSTEFVSLASKYVDALKYGVVF 133

DB 62 LLVWHLFAVRNAPVQKLYGYRIANTQVEAVENSTTEFPDLISKVISTRTLYNGKD 121

QY 134 LGPYHKESAVTAFSEGS--VIAYYWSFSLPQHLVEAEERVMAEERVVMLPPRARSIKS 190

DB 122 IAPLYQCSISAFSEGSNNVGVYWSFVPAFEAEFEKALISE--LKLPTVNLQRA 178

QY 191 FVTVSVVAFPTDQKTVQTONSCSPGLHARGVELMRTTGPPSPYPAHARQOMLRG 250

DB 179 FAVSVLVAFPTDQKTVQTONSCSPGLHARGVELMRTTGPPSPYPAHARQOMLRG 238

QY 251 DADVSLTTFPSFDLASCDEGSDLVTVYNTLSPMEPHALVOLCTGYPPSYNTLTFHSSON 310

DB 239 DAGVMIRKFTFKMEKCANAGDPVWVYDLSLPIEPAAQRLCGIYPPSYNTLTFHSSON 298

QY 311 VILITLITNERRHHPGEATFOLFPRMSSCGGRRLKXQGTENSPYPYGHYPNIDCTMNI 370

DB 299 VMLTTLVTDVNVCKPGLAFSPFPKTSLCGGYTRDASGVFTSPFPGHPPKIECIWDI 358

QY 371 EVPNQHVAVRPFKFFYLLPEGVPAAGTCPKDQVVEINGEKSGERSQFVYVTSNKNITVRFH 430

DB 359 QVPDNKFKVLRFMNFYLAEPGVPTKCTKDQFVEINGQYGERKFFVVSNNSSKMSRVFV 418

QY 431 SDQSYDTGTFLAELYSIDSSDPCPGQFTCTGRCIRKELACDQWADCTDSDELACSCDA 490

DB 419 SDQSYDTGTFLAELYSIEPPNCPDQFACKSGHCIRLDQKCDQWADCTDSDELACSCCTA 478

QY 491 GHOFTCKN-KECKPLFMVQCSVNDCCGNNSDEGSCSPAQFPRCSNGKCLSKQOQNGKDD 549

DB 479 -LQPRCTNSKLCXRSYVCDGVNDCCGSSDELACQCPNNITKYCGNGKCLIPESQCDRTDN 537

QY 550 CGDSDASCPKVVVYCTGHTYRCLNGLCISKNGPECKDEKDCSDSDSDE--KDCDGLR 607

DB 538 CGDSDASCPKVVVYCTGHTYRCLNGLCISKNGPECKDEKDCSDSDSDE--KDCDGLR 597

QY 608 SFTROARVVGCTDADDEGEWPMQVSLHAGGCHICGASLISPNMLVSAHICYIDRGRRYS 667

DB 598 PFTKSRIVGVANA----- 611

QY 668 DPTQWTFGLGHDSQSRASRGVQERLRKRIISHPFNFDPFDYDIALBLEKPAEYSSWV 727

DB 612 ----- 611

QY 728 RPICLPDASHVFPAGKAIWYTGNGHTQYGGTALLQKGEIRVINYQTCENLLPQQTTPR 787

DB 612 ----- 611

QY 788 MCVGLSLSGVSDCGDSCGPLSSVEADGRIFGAGVYSWGDGCAGRKKPVYTRLPFRD 847

DB 612 -----DSGGLPSLSEVLEINRYLAGIVSWGEGCARPKRPGVYTRVAMMRD 655

QY 848 WIKENTGV 855

DB 656 WIKENTGV 663

RESULT 9

Q6PP94 PRELIMINARY; PRT; 799 AA.

AC Q6PP94; 05-JUL-2004 (TREMblrel. 27, Created)

DT 05-JUL-2004 (TREMblrel. 27, Last sequence update)

DT 05-JUL-2004 (TREMblrel. 27, Last annotation update)

DE Tmprs6 protein.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=FVB/N; TISSUE=liver;

RX MEDLINE=22388257; PubMed=12477932;

RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heieh F., Ditschenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Uedlin T.B., Toshiyuki S., Carrinci P., Prange C., Raha S.S., Loguelli N.A., Peters G.J., Abramson R.D., Mullaly S.J., Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A., Whiteley M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,

RA Krzywinski M.I., Skalska U., Smalius D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Liver;
RA Strausberg R.;
RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Belongs to peptidase family 11.
DR EMBL: BC057674; AAH57674.1; -
DR GO: GO:0005886; C:plasma membrane; IDA.
DR InterPro: IPR000859; CUB.
DR InterPro: IPR002172; LDL_receptor_A.
DR InterPro: IPR001254; Peptidase_S1-
DR InterPro: IPR003003; Pept_Ser_Cys.
DR Pfam: PF00089; Trypsin_1.
DR Pfam: PF00057; Ldl_recept_a; 2.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR PRINTS: PR00261; LDLRECEPTOR.
DR SMART: SM00192; LDLa; 3.
DR SMART: SM00202; Tryp_Spc; 1.
DR PROSITE: PS01180; CUB; 1.
DR PROSITE: PS01209; LDLRA_1; 1.
DR PROSITE: PS50068; LDLRA_2; 3.
DR PROSITE: PS50240; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
KW Hydrolyase; Protease; Serine protease.
SQ SEQUENCE 799 AA; 89557 MW; 16315A646A4D5288 CRC64;

Query Match 23.8%; Score 114.5; DB 2; Length 799;
Best Local Similarity 31.6%; Pred. No. 2,1e-68;
Matches 266; Conservative 140; Mismatches 347; Indels 90; Gaps 25;

QY 33 EEVVE-----FLPVNVKVKKEKGGKGRVWVLAVALIGLLVLLGIGPLVMH-LQYR-DVRV 86
DB 20 EEAEPBGKFKFPRNTRK-----KNRDVVRTPPLLVLAALVSAAGWMLWYLGKAEVTV 74
QY 87 QKVNKGKMRITNEFVAVAYENSNTSEFVSLASKYKDALKLYSVPLGPRPKESANTAF 146
DB 75 SQVYSGSLRVLNRHSQDLGRREIARSESAKQKMLQELVASTR-LGTYNSSVYSF 133
QY 147 SEGSVIAYVSESEFIPQHLVEAEARVAAEEVV--MLPPRARSLSKSPVTSVVAEPDTSK 204
DB 134 GEGPLTGFPMFLIDIPY-----QRLTSPSEVARELVDELSSSTLASIKTEYVDPE 188
QY 205 --TVQRTQDNSCSFGLHARGVELMRFTTPG--FP-DSPYPAHARQWALRGDADSVLSLT 259
DB 189 GLVILEASVNDIVIANSTLGCYRYSVNPQGVLPKGPDDQTTSCLMHLLQGPEDLMIKVR 248
QY 260 FRSDLASCDERGGDLTVVNTLSPMEPHALVQCGTYPPSYNTLTHSSQVLLITLTN 319
DB 249 L-ENTRVDCRDR---VAAMDAGPLEKRLTLTYGSGROBPVNEVLAAGSVAWVWKKG 303
QY 320 TERHPPGF---EATFPGLPRMS--SCGRLKAKGTFSPPYGHVPPNIDCTNIEVFN 374
DB 304 MHSYIDFLLSVKSAFQDCQVNLTEGRLL-DTQGLPRLTPYPSYSPSTHCSNHLTVPS 362
QY 375 NQHVKVFKEFFYLEPVGPACTCPDYVEINGEKYCGER-----SQFVYVSNKKTIVRF 429
DB 363 LDYGLALMFDAVYALRQKYNRLCTQGGWMIQNRLLCGFRLLQPYAERLPMVAASGVITINF 422
QY 430 HSDDSYIDTGLFAYLISYDSDPFGPGFTCTGTGCIKEKLRCQDMACTDHSDELNSCD 489
DB 423 TSQISLTLGPQGVYVSLYNOSDPCGGEFLCSV----- 454
QY 490 AGHQFTCKNKKCKFLFWVCDVNDGNSDEGSCSPAQTRFC-SNGCKLSKSQCGNGKD 548
DB 455 -----NGLCVP---ACDGIKDCPNGLDERNCVCRAM-FQCGEDSTCTISLPRVCDROP 502

QY 549 DCGDSDSEASCPKRVNVVTKHTYRCLMGLCLSKNGPECDGKEDSGSDGKDCGLRS 608
DB 503 DCLNGSDDEQCQR--GVPCGCTFFPGCEDRSCKVKKPBCQGDGDCDSDGDCGLQG 560
QY 609 FTQDARVVGTTDDBEEMPVQVSLHLAGCHTGAASLISNMVSAHCVITDRGRFYSD 668
DB 561 LS--SRIVGGTVESEGEWPAQLQIRGR-HICGGLADRWVITAAHCEQD--SMAS 614
QY 669 PTQMTAFILGHDSORSAPVQERRLKRIISHPPFNDFTFYDIALLEKPAEVSMMVR 728
DB 615 PKMTVFLGMRQNSR-WPEGVSFYSRLFLPHYHEDSHDYVALLQDHPVYSATVR 673
QY 729 PCLPDAASHVFPAGKAIWTTGWHTGYGTGALLLOKGLIRVYNOTTCENLPLPQOTPRM 788
DB 674 PVCLPASHFFPEPGCHCWITGMCAGOREGFSVNTLQKVQVLPDOLCSAEVRYQVSPRM 733
QY 789 MCYGFSLSGVDSQCGSGGSLVSEADGRFGSGVGVSWGDCGARRKPGYTRPLPRDW 848
DB 734 LCAGYRKGRKDCQSDSGGLVCRPPSGRWFGLVSGWGLGCGRPNFPGYTVTRVINY 793
QY 849 IKE 851
DB 794 IQQ 796

RESULT 10
AAH57674 PRELIMINARY; PRT; 799 AA.
ID AAH57674
AC AAH57674
DT 02-MAR-2004 (TEMBLrel. 27, Created)
DT 02-MAR-2004 (TEMBLrel. 27, Last sequence update)
DT 02-MAR-2004 (TEMBLrel. 27, Last annotation update)
DE Trpase6 protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Liver;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marsala K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mulhany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Pabey J., Helton B., Kettelman M., Madan A.C., Shevchenko Y., Bouffard G.G.,
RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalius D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Liver;
RA Strausberg R.;
RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC057674; AAH57674.1; -
SQ SEQUENCE 799 AA; 89557 MW; 16315A646A4D5288 CRC64;

Query Match 23.8%; Score 114.5; DB 2; Length 799;
Best Local Similarity 31.6%; Pred. No. 2,1e-68;
Matches 266; Conservative 140; Mismatches 347; Indels 90; Gaps 25;

33 BEGVE-----FLPVNNVKVKGKGGQRMVYLAAILGLLVLLGIGLVWH-LOYR-DVRY 86
 20 EEAABPECKFKPPKTKR-----KRDYVRYFTPLILVALVSAQWMLWTLGKAEYTV 74
 87 QKKNQYRITNENFVDAVENSSTEFVSLASKYVDALKLYSQVFLGPKHESAVYAF 146
 75 SQVYSGSLRVNRFPSODLGRRESIAFRSBSAKQKMLQELVASTR-LGYTSSSVYSF 133
 147 SEGVIAYVYSEFSIPQHLVEAEERVAERVY-MLPPRARSLKSFVYTVVAPFPD 204
 134 GEGPLTFEFPILIPY-----ORLTLSPEVVELLDDELISNSTLASXKTEEVDP 188
 205 --TVQRQDNCSQSPGLHARGVELMRFTTGR--FP-DSYPHARQWALRGDADSVLSLT 259
 189 GLVILASVNDIVVNLSTLGCYRSYVNPQVLPKGDQDTTCLMLQCPEDMLTVR 248
 260 FRSPDLASCDERGDLYVYVNTLSPMEHALVOLCGTYPSPYNTLPHSSQVLLITLTN 319
 249 L-EWTRVDCRDR-----VAMYDAAGPLEKRLITSVYCGRQBPMEVLASGVMVWVKG 303
 320 TERHHPGF-----EATFPOLPMS--SCGRLKAGCTSPYFGHYPPNIDCTWNIEVPN 374
 304 MHSYDPLFLSVKSAFQDCQVNLTLGRL-DTQGFRTPYPSYSPSTHCSMHLTVPS 362
 375 NQHVKVPKFFYLLPEGVPACTCPKDYVEINGEKYCGR-----SQFVYTSNKKITVR 429
 363 LDYGLALWFDAVYALRQKTRNLCTQGMIONRRLCGRTTQPAERLPMVASDVTINF 422
 430 HSDSYDTDTGLAEYLSYDSSDPGCGQPTCTGRCIRKELKCDGADCTHSDILNCS 489
 423 TSQSLTGPQVQVYVYSLYNSDPCGEGFLSV----- 454
 490 AGHOFCKKCKKCKLFWVCDVNDGNSDEQCSGCPAQTRC-SNGCKLSKQCCNKD 548
 455 -----NGLCVP---ACDGIKCCPGLDERNCCRAM-FOCEDSTCISLPRVCQRP 502
 549 DCGSGSDASCPKVVVYCTHATYRCLNGLCLSKNPNCDKEDCDSDSDKDCDGLRS 608
 503 DCLNGSDEBQCF--GVPCGFTTQCEBRSVCKKPNPCDQSDSDSDSDHCHDCGLG 560
 609 FTROAVVGTDADEGEMPMQVSLHALQGHICGASLISPMWLVSAAHYCIDDRGFRYS 668
 561 LS--SRITGTVSSGEMPMQASIQIRR-HICGALADRWITAAHCFQD--SWAS 614
 669 PTQWTAFLGLHDGORSAPVQERLKAIIISHPPNFTFYDIALLEKPAEYSSNR 728
 615 PKLWTVFLGKMRQNSR-WPGEVSFVSRFLPLHPYHEBDSHYDVALLQDHPVYSATVR 673
 729 PICLPDASHVPAGKAVVTGSMHTQVGTGALLIQKEIRIVNQTGENTLPOQTPRM 788
 674 PVLCPANSHFEPGCHWITGMAQREGPVSNLQKVDVLPQDLCEAARYQVSPRM 733
 789 MCVFSLSGVSDSCGDSGGLPSVEADRIFGAGVSWGDCAGANKKPYVTRLPFLRDM 848
 734 LCAVYRKGKXKACGDSGGLPVCREPSRWFLAGLVSGLCGGRNPFVYTRVAVNW 793
 849 IKE 851
 794 IQQ 796

NCBI_Taxid=10090;
 [1]
 SEQUENCE FROM N.A. AND TISSUE SPECIFICITY.
 STRAIN=C57BL/6J;
 MEDLINE=22755759; PubMed=12744720; DOI=10.1042/BJ20030390;
 Hooper J.D., Campagnolo L., Goodarzi G., Truong T.N., Stuhlmann H.,
 Outgley J.P.;
 "Mouse matrilase-2: identification, characterization and comparative
 mRNA expression analysis with mouse hepsin in adult and embryonic
 tissues." J. 373:689-702(2003).
 [2]
 SEQUENCE FROM N.A.
 STRAIN=C57BL/6J; TISSUE=Liver;
 MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;
 Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
 Nikaio I., Oseko N., Saito R., Suzuki H., Yamanaoka I., Kiyosawa H.,
 Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gotohori T.,
 Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
 Schiraldi L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
 Blake J.A., Brad D., Brusic V., Chochia C., Corbett L.E., Cousins S.,
 Dalla B., Dragani T.A., Fletcher C.P., Forrest A., Frazer K.S.,
 Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
 Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
 Kanai A., Kawai H., Kawasawa Y., Kedzierski R.M., King B.L.,
 Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
 Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
 Nagashima T., Numata K., Okido T., Pavan W.J., Petter G., Pesole G.,
 Petrovsky N., Pillai R., Pontius J.V., Qi D., Ramachandran S.,
 Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
 Sadelain A., Schneider C., Seiple C.A., Setou M., Shimada K.,
 Sultana R., Takeda Y., Taylor M.S., Teasdale R.D., Tomita M.,
 Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
 Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
 Yuan Z., Zavolan M., Zhu Y., Zimmer A., Cernici P., Hayatsu N.,
 Hirozane-Kishikawa T., Kono H., Nakamura M., Sakazume N., Sato K.,
 Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
 Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
 Miyazaki A., Sakai K., Sasaki D., Shibata K., Shingawa A.,
 Yamanishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
 Birney E., Hayashizaki Y.;
 "Analysis of the mouse transcriptome based on functional annotation of
 60,770 full-length cDNAs." J.
 Nature 420:563-573(2002).
 [3]
 SEQUENCE FROM N.A.
 TISSUE=Liver;
 MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D.,
 Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heien F.,
 DiCicco L., Marisina K., Farmer A.A., Rubin G.M., Hong L.,
 Stappleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 Brownstein M.J., Udell T.B., Toshitsuki S., Giannini P., Prange C.,
 Rana S.S., Loguelfano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 Bosak S.A., McEwan P.U., McKernan K.U., Malek J.A., Guneratne P.H.,
 Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huliy S.W.,
 Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 Fahney J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Skalska D.E.,
 Scherch A., Schein J.B., Jones S.J.M., Matra M.A.;
 "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences." J.
 Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 [4]
 REVIEW.
 MEDLINE=22668120; PubMed=12784999;
 Netzel-Arnett S., Hooper J.D., Szabo R., Madson E.L., Outgley J.P.,
 Bugge T.H., Antalis T.M.;

RT "Membrane anchored serine proteases: a rapidly expanding group of cell
RT surface proteolytic enzymes with potential roles in cancer.";
RL Cancer Metastasis Rev. 22:237-258 (2003).
CC -1- FUNCTION: May play a specialized role in matrix remodeling
CC processes in liver (By similarity).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein.
CC -1- TISSUE SPECIFICITY: Expressed at highest levels in liver, kidney
CC and uterus.
CC -1- DEVELOPMENTAL STAGE: Expressed at higher levels from 12.5 dpc to
CC 15.5 dpc with a peak at 13.5 dpc. Expression in the developing
CC liver as well as a restricted set of embryonic epithelial cells of
CC the nasal cavity and pharyngo-lymphatic tubes.
CC -1- SIMILARITY: Belongs to peptidase family S1.
CC -1- SIMILARITY: Contains 3 LDL-receptor class A domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AY240929; AAP69827.1; -
CC EMBL; AK004939; BAB33684.2; -
CC EMBL; BC029645; AAH29645.2; -
CC HSSP; P00760; 1E2X.
CC MEROPS: S01.308; -
CC MGD; MGI:1919003; Tmpres6.
CC GO; GO:0016021; C:Integral to membrane; ISS.
CC GO; GO:0005886; C:Plasma membrane; IDA.
CC GO; GO:0004252; F:serine-type endopeptidase activity; ISS.
CC GO; GO:0001525; P:angiogenesis; ISS.
CC GO; GO:0030198; P:extracellular matrix organization and bioge. . .; ISS.
CC GO; GO:0007242; P:fibrinolysis; ISS.
CC GO; GO:0006508; P:intracellular signaling cascade; ISS.
CC GO; GO:0006508; P:proteolysis and peptidolysis; ISS.
CC InterPro; IPR000859; CUB.
CC InterPro; IPR002172; LDL_receptor A.
CC InterPro; IPR001254; Peptidase_S1.
CC InterPro; IPR001314; Peptidase_S1A.
CC InterPro; IPR009003; Pept_Ser_Cys.
CC Pfam; PF00057; Ldl_recept_a; 2.
CC PRINTS; PR00089; Trypsin; 1.
CC PRINTS; PR00722; Chymotrypsin.
CC PRINTS; PR00261; LDLRECEPTOR.
CC PROSITE; PS01180; CUB; 1.
CC PROSITE; PS01209; LDLR_1; 1.
CC PROSITE; PS50068; LDLR_2; 3.
CC PROSITE; PS50240; TRYPSIN_DOM; 1.
CC PROSITE; PS00134; TRYPSIN_HIS; 1.
CC PROSITE; PS00135; TRYPSIN_SER; 1.
KW Glycoprotein; Hydrolase; Repeat; Serine protease; Signal-anchor;
KW Transmembrane.
FT DOMAIN 1 59 Cytoplasmic (Potential).
FT TRANSSEM 60 80 Signal-anchor for type II membrane
FT 81 811 protein (Potential).
FT DOMAIN 213 336 Extracellular (Potential).
FT DOMAIN 323 440 CUB 1.
FT DOMAIN 445 477 CUB 2.
FT DOMAIN 478 514 LDL-receptor class A 1.
FT DOMAIN 518 555 LDL-receptor class A 2.
FT DOMAIN 565 799 LDL-receptor class A 3.
FT ACT_SITE 617 799 Serine protease.
FT ACT_SITE 668 668 Charge relay system (By similarity).
FT ACT_SITE 668 668 Charge relay system (By similarity).
FT ACT_SITE 762 762 Charge relay system (By similarity).
FT CARBOHYD 138 138 N-linked (GlcNAc. . .) (Potential).
FT CARBOHYD 184 184 N-linked (GlcNAc. . .) (Potential).
FT CARBOHYD 216 216 N-linked (GlcNAc. . .) (Potential).
FT CARBOHYD 338 338 N-linked (GlcNAc. . .) (Potential).
FT CARBOHYD 433 433 N-linked (GlcNAc. . .) (Potential).

FT CARBOHYD 453 453 N-linked (GlcNAc. . .) (Potential).
FT CARBOHYD 518 518 N-linked (GlcNAc. . .) (Potential).
FT CONFLICT 690 690 P -> PP (in Ref. 2).
SQ SEQUENCE 811 AA; 90978 MW; 32EB3E7C3127801B CRC64;
Query Match 23.8%; Score 1114.5; DB 1; Length 811;
Best Local Similarity 31.6%; Pred. No. 2; Le-68;
Matches 266; Conservative 140; Mismatches 347; Indels 90; Gaps 25;
Qy 33 BEGVE-----PLPANNVKKVKGPRGVVLAALVGLVLTGLGVLMH-LQYR-DVRY 86
Dy 32 BEAAEPGKPKPKTKR-----KARDYRFTLLVLAALVAGVLAALVFLGKAEVY 86
Qy 87 QYKNGYRITNENPVDAENSNSTFVSLASVKAALKLYSGVPLGPKYKESAVTAF 146
Dy 87 QYVSGSLRVLNHFSSODLRRRESIAFRSESAKAQKMLQELVASTR-LGYVNSSVYSYF 145
Qy 147 SEGSVLAIVYSESIQHLVBEAEVMAERVY-NLPRAKLKSVTVTSVVAFPDSD 204
Dy 146 GEGPLTCFFWFLIDIPY-----QRLTSPREVRELLVDELLSNSTLASKYTEYVDPE 200
Qy 205 --TVQRTQDSCSFGHARGVEIMRFTTPG--PP-DSPYPAHARCOWALRGADSVSLT 259
Dy 201 GLVILEASVNDIVYVNLSTLCYRYSYVNEGQVLPKGPDDQTSSCLMHLGGPDDMLTKR 260
Qy 260 FRSFDLASCDERGSDLVTYVNTLSPEPHALVOLCGTTPSYNLTPHSQVLLITLTN 319
Dy 261 L-EWTRVDCRDR-----VAMTDAAGPLEKRLITSVGCSSRQEPVMEVLAASVAAVWVKG 315
Qy 320 TERRHRPF-----EATFQPLRMS-SCGGRARKQGTNSPYVGHYPNIDCMNIEVPN 374
Dy 316 MHSYVPEFLSVSVAFQDCQVNLTEGR-L-DTQGLFRLTYPSYSPSHCSMHLTPS 374
Qy 375 NOHVYKAFKFFVLEBGPVAPGCPKDYVEINGEKYGER-----SQFVYVSNKKTIVRF 429
Dy 375 LDYGLAMPAYVLAIRKQXNRLCTQGGWMIQNRLGFRLLQYARIRPVADGVITNF 424
Qy 430 HSDQSYTDGFLAEVLSYDSSDPCQGFCTGRCIRKELCDGMADCTDHSDELNCSD 489
Dy 435 TSQISLTLPQGVQYYSLYNQSDPCPGEFLLCSV----- 466
Qy 490 AGHQFTCKNFKCPRLFWVCDVNDGDNDSQSCSPACQFRFC-SNGCKLSKQCGNKGD 548
Dy 467 -----NGLCVP---ACDGIQDCPNGLDERNCVCRA-M-FQCEQDSTCISLPRVCDROP 514
Qy 549 DCGDGSDEASCPYVNVYCTKHTYRCINGLCSKGNPCDCKDSCDSDKDCDGLRS 608
Dy 515 DCLANGSDEBQCE--GVPCSTFTFCEDRSQVKKPNECDGGSDGSDGHCDCGLQG 572
Qy 609 FTRQARVVGTDADDEGEWPMQVSLHALGQGHICGASLISPNMLVSAACHYIDRGRFYSD 668
Dy 573 LS--SRIVGTVSSEGEWPMQASLQIRGR-HICGALLADRWITTAHACQCEQ---SMAS 626
Qy 669 PTQWTAFLGHDQSQASAPVQGERLKRISHSFPNDFPTDYVILLLEKPAEVSQMR 728
Dy 627 PKLMTVFLGKQRNSR-WPEGEVSFKVSRLLPHYHEDSDHYVALLQDHPVYSATVR 665
Qy 729 PICLPDASHFPAKAIWYGMGHTQVGTGALLQKGEIRVINTQTCENLFPQITPRM 788
Dy 686 PVLCLPARSHFPEQGHCHWIKWAGRGEGPVSTLQKDVYQPLQDLSAAYVQVSPRM 745
Qy 789 MCVGFSLSGVDSCQDSSGGPLSSVEADGRIFGAGVSWMDGCAGRNKPGVYTRLPFRDW 848
Dy 746 LCAGYRKGKDAQCGDSGGLVCRBPSGRWFLAGVSWGLGCGRPNPFVGYTRVTVIWM 805
Qy 849 IKX 851
Dy 806 IQQ 808
RESULT 12
TMS6_HUMAN
ID TMS6_HUMAN STANDARD; PRT; 811 AA.

AC O81U80; O81U82; O81XV8;
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Transmembrane protease, serine 6 (EC 3.4.21.-) (Matricase-2).
 GN Name=TMPRSS6;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RA Hooper J.D., Quigley J.P.;
 RT "TMPRSS6, a new type II transmembrane serine protease";
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20057165; PubMed=10591208; DOI=10.1038/990031;
 RA Dunham I., Hunt A.R., Collins J.E., Brunkiewicz R., Beare D.M.,
 RA Clamp M., Smink L.J., Atencough R., Almeida J.P., Babbage A.K.,
 RA Begganey C., Bailey U., Barlow K.F., Bates K.N., Beasley O.P.,
 RA Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,
 RA Buttrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,
 RA Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,
 RA Conway D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,
 RA Dhami P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,
 RA Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,
 RA Gilbert J.G., Goward M.E., Graham D.V., Griffiths M.N.D., Hall C.,
 RA Hall R.E., Hall-Tamlyn G., Heathcote R.M., Ho S., Holmes S.,
 RA Hunt S.E., Jones M.C., Kersey J., Kimberley A.M., King A.,
 RA Laird G.K., Langford C.F., Leverhulme M.A., Lloyd C., Lloyd D.M.,
 RA Martyn I.D., Maehrighi-Mohammadi M., Matthews L.H., Mccann O.T.,
 RA Mcclay J., McLaren S., McMurtry A.A., Milne S.A., Mortimore B.J.,
 RA Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,
 RA Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L.,
 RA Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,
 RA Soderlund C., Spragon L., Steward C.A., Sulston J.E., Swann R.M.,
 RA Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,
 RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilting L.,
 RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,
 RA Minoshima S., Kawaaki K., Sasaki T., Aakawa S., Kudoh J.,
 RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyama S.,
 RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,
 RA Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,
 RA Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Pan H.,
 RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shauli S., Sloan D., Song L.,
 RA Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,
 RA Zhan M., Zhang G., Chisoe S., Murray J., Miller N., Minx P.,
 RA Fulton R., Johnson D., Bems G., Bentley D., Bradshaw H., Bourne S.,
 RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,
 RA Hinds K., Kemp K., Latreille P., Layman D., Ozeraky P., Rohlfing T.,
 RA Scheet P., Walker K., Wamsley A., Wohlmann P., Pepin K., Nelson J.,
 RA Korf I., Bedell J.A., Hillier L.W., Mardis E., Waterston R.,
 RA Wilson R., Emanuel B.S., Shaikh T., Kurahashi H., Salter S.,
 RA Budarf M.L., Mcdermid H.E., Johnson A., Wong A.C., Morrow B.E.,
 RA Edelman L., Kim U.J., Shizuya H., Simon M.I., Duncan J.P.,
 RA Peyrard M., Kedia D., Seroussi E., Franssen I., Tapia I., Bruder C.E.,
 RA O'Brien K.P., Wilkinson P., Bodenteich A., Hartman K., Hu X.,
 RA Khan A.S., Lane L., Tiliakou Y., Wright H.;
 RT "The DNA sequence of human chromosome 22";
 RL Nature 402:489-495(1999).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Brain;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Straube R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner K.H., Schaefer C.F., Bhat N.K.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Wax S.I., Wang J., Hsieh F.,
 RA Datschenko L., Maruina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ueda T.B., Toshitoki S., Carinci P., Prange C.,
 RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mulvihy S.J.,

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Rahey U., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bonfield G.G.,
 RA Blakeley R.W., Touchman A.J., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Buttefield Y.S.N., Krzywinski M.I., Skalka U., Small D.E.,
 RA Scherach A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [4]
 RP SEQUENCE OF 9-811 FROM N.A. (ISOFORM 1), AND TISSUE SPECIFICITY.
 RC TISSUE=Testis;
 RX MEDLINE=2241917; PubMed=12149247; DOI=10.1074/jbc.M203007200;
 RA Velaasco G., Cal S., Quesada V., Sanchez L.M., Lopez-Otin C.;
 RT "Matricase-2, a membrane-bound mosaic serine proteinase predominantly
 RT expressed in human liver and showing degrading activity against
 RT extracellular matrix proteins";
 RL J. Biol. Chem. 277:37637-37646(2002).
 RN [5]
 RP REVIEW.
 RX MEDLINE=2268120; PubMed=12784999;
 RA Netzel-Annett S., Hooper J.D., Szabo R., Madison E.L., Quigley J.P.,
 RA Bugge T.H., Anstall T.M.;
 RT "Membrane anchored serine proteases: a rapidly expanding group of cell
 RT surface proteolytic enzymes with potential roles in cancer";
 RL Cancer Metastasis Rev. 22:237-258(2003).
 CC -1- FUNCTION: May play a specialized role in matrix remodeling
 CC processes in liver.
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=O81U80-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=O81U80-2; Sequence=VSP 008379, VSP 008380;
 CC Note=No experimental confirmation available;
 CC -1- TISSUE SPECIFICITY: Liver specific.
 CC -1- SIMILARITY: Belongs to peptidase family 11.
 CC -1- SIMILARITY: Contains 2 CUB domains.
 CC -1- SIMILARITY: Contains 3 LDL-receptor class A domains.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AJ319876; CAC85953.1; ALT_INT.
 CC EMBL; AY055383; AAL16413.1;
 CC EMBL; AY055384; AAL16414.1;
 CC EMBL; AL022314; NOT ANNOTATED_CDS.
 CC EMBL; BC039082; AAH39082.1;
 CC HSSP; P00760; 1BZX.
 CC Genes; HGNC:16517; TMPRSS6.
 CC InterPro; IPR000859; CUB.
 CC InterPro; IPR002172; LDL_receptor_A.
 CC InterPro; IPR001254; Peptidase_S1.
 CC InterPro; IPR009003; Pept_ser_Cys.
 CC Pfam; PF00431; CUB; 1.
 CC Pfam; PF00057; LDL_recept_a; 2.
 CC Pfam; PF00089; Trypsin; 1.
 CC PRINTS; PR00722; CHYMOTRYPSIN.
 CC PRINTS; PR00261; LDLRECEPTOR.
 CC SMART; SM00042; CUB; 1.
 CC SMART; SM00192; LDLa; 3.
 CC SMART; SM00020; TRYP_SPC; 1.
 CC PROSITE; PS01180; CUB; 1.

Db 464 ---ACDGVKCCPNGLDERNCVCRA-TFOCKEDSTCISLPKVCODQPPCLNGSDEEOCE- 518
Qy 563 NVVTCRHTYRCNLGLCLSKGNPCDCKEDCSDGSDKDCGLRSRTROARVVGTDAD 622
Db 519 -GVPCGFTTFOCEDRSCVCKKPNQCDRPPDRSDSDHDCDGLQG--PSSRTVGA VSS 575
Qy 623 EGEWPMQVSLHALGQGHICGASLISPMWLVSAAHCYIDRGRFYSDPTOWTAFLGHDOS 682
Db 576 EGEWPMQVSLHALGQGHICGASLISPMWLVSAAHCYIDRGRFYSDPTOWTAFLGHDOS 682
Qy 683 QRSAPGVQERLRKRIISHPPFNDFPYDIALLEKRPAYSSMVRPCLPASHVPAG 742
Db 632 SR-WPGGVSPKVSRLIHPYHEBDSDYVALQLDHPVRSAAVRPVCCLPARSHFPEPG 690
Qy 743 KAIWVTGMHTOYGGTGA-LIQQEIRVINGTTCENTLPPQITPRMMCVGFLSGVNSCQ 802
Db 691 LHCHITGMALRREGGPIISNALQKVVDLIPDOLCEA YRYQVTPRMLCAGYRKCKDQACQ 750
Qy 803 GDSGGLPSVYADGRIFGAGVSWGDCAGRNKPGVYTRPLFRDWIKX 851
Db 751 GDSGGLPVCKALSGRWFLAGLVSGLGCGRPNTYGVYTRITGVISWIOQ 799

RESULT 15

Q61CC2 PRELIMINARY; PRT; 824 AA.
ID 061CC2
AC 061CC2;
DT 05-JUL-2004 (Tremblrel. 27, Created)
DT 05-JUL-2004 (Tremblrel. 27, Last sequence update)
DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)
DE D11170K4.2 protein.
GN Name=d11170K4.2;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Collins J.E., Wright C.L., Edwards C.A., Davis M.P., Grinham J.A.,
RA Cole C.G., Goward M.E., Aguado B., Mallya M., Mokrab Y., Huckle E.J.,
RA Beare D.M., Dunham I.,
RL Submitted (May-2004) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Belongs to peptidase family S1.
DR EMBL; CR456446; CAG30332.1; -.
DR InterPro; IPR000859; CUB.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR InterPro; IPR009003; Pept_Ser_Cys.
DR Pfam; PF00431; CUB; 1.
DR Pfam; PF00057; LDL_recept_a; 2.
DR Pfam; PF00089; Trypsin; 1.
DR PRINTS; PRO0722; CHYMOTRYPSIN.
DR PRINTS; PRO0261; LDLRECEPTOR.
DR SMART; SM00192; LDla; 3.
DR SMART; SM00202; TRYP_SPC; 1.
DR PROSITE; PS01180; CUB; 1.
DR PROSITE; PS01209; LDLRA_1; 1.
DR PROSITE; PS50068; LDLRA_2; 3.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KM Hydrolyase; Protease; Serine protease.
SQ SEQUENCE 824 AA; 91333 MW; A74F186406041F7B CRC64;

Query Match 23.3%; Score 1092; DB 2; Length 824;

Best Local Similarity 30.3%; Pred. No. 7.8e-67;

Matches 276; Conservative 135; Mismatches 330; Indels 170; Gaps 29;
Qy 11 GGPDPFGAGLYNNRHHKNGLEBGFPLPVNNVKKYKEKGGPGWVYLAVALIGLLVYL 70
Db 11 GGQGGGSDG-----EAAEPGGMFYACE-----DSKRKANGYRLVPLFVYL--ALLVLA 57

Qy 71 GIGFLVWH-LQYR-DYRVQKKNYKRIINENFVDAEYNSNSTEYVSLASKVDALKLY 128
Db 58 SAGVLLMFLYGYAEVWVSVVSGSLRVLNRRHSQDLTRRESAFAFSEPAKQMKELI 117
Qy 129 SGVPFLGPNHKEAAVYAFSEGSYIAYYSEFSIPQULVEAEVYAEVYVMLPPARSL 188
Db 118 TSTR-LGYTNSSSVYSFGEGPLTCFFWTLQIPFH-----RRLMSPEV--V 162
Qy 189 KSFVYVSVAFPDSTKYVGTODNSCSFGHARGVLEMF-----TTPGFPDSPY- 238
Db 163 QALIVEELSTVNSSAAVPRAB-----YEVDPGLVILASVYDIAANSTLGCTYSYV 218
Qy 239 -----PAH--ARCGMALRGDADSVLSLTFRSFDLASCDERSDLYVNTLSPEM 286
Db 219 GQGVRLRKGPDLHASSCLMHLQGPDLMLKRL--EWTLAECDR-----LAMYDVAGPLE 273
Qy 287 PHALVOL--CGTPPSYNTL-----FHSSQVLLITLTNTERRRPGREATP 331
Db 274 KRLITVYGCGRQEPVEVYLAGAIVAVWKGLHSYBPFVLSV-----QPVV 332
Qy 332 FQLPMSGCGRLKKAQGTENSPYRPGHYRPNIDCTWNIENVNNOHVYKFKFYLLERPG 391
Db 323 FQACEVNLIDNRLDSQVILSTYFPYSYSPQTHCSMHLTVESLDYGLALMPAYALRRQ 382
Qy 392 VPAGTCQVVEINERKYCG-----ERSQFVYVTSNKNITVRPHSDSYTDTGFLAE 443
Db 383 KYDLPTQGGWNTIQNRLCGRLILOPYAERIPVATAG--ITINFTSQSLTGPGVRVH 439
Qy 444 YLSYDSDPCEPGQFTGRTGRTKELKRCQGMADCTHSDIELNCSQAHQFTCKNFCXP 503
Db 440 YGLYNSDDPCGSEFLSV-----NGLCYE 463
Qy 504 LFWVCSVNDGCGNSDEGCSCPAOTFRCSNGKCLSKSQCCNGKXKDCDGDSDSEASCPTY 562
Db 464 ---ACDGVKDCRPNGLDERNCVCRA-TFOCKEDSTCISLPKVCODQPPCLNGSDEEOCE- 518
Qy 563 NVVTCRHTYRCNLGLCLSKGNPCDCKEDCSDGSDKDCGLRSRTROARVVGTDAD 622
Db 519 -GVPCGFTTFOCEDRSCVCKKPNQCDRPPDRSDSDHDCDGLQG--PSSRTVGA VSS 575
Qy 623 EGEWPMQVSLHALGQGHICGASLISPMWLVSAAHCYIDRGRFYSDPTOWTAFLGHDOS 682
Db 576 EGEWPMQVSLHALGQGHICGASLISPMWLVSAAHCYIDRGRFYSDPTOWTAFLGHDOS 682
Qy 683 QRSAPGVQERLRKRIISHPPFNDFPYDIALLEKRPAYSSMVRPCLPASHVPAG 742
Db 632 SR-WPGGVSPKVSRLIHPYHEBDSDYVALQLDHPVRSAAVRPVCCLPARSHFPEPG 690
Qy 743 KAIWVTGMHTOYGGTGA-----LIIQGEIRVINGTTCENTL 780
Db 691 LHCHITGMALRREGALRADVALFYGMRNQSESTCCPISNALQKVVDLIPDOLCEAY 750
Qy 781 PQITPRMVCVGLSGVNSCGDSDGGLSVADGRIFGAGVSWGDCAGRNKPGVYTR 840
Db 751 RYGVTPRMLCAGYRKCKDQACQDSGGLPVCKALSGRWFLAGLVSGLGCGRPNTYGVY 810
Qy 841 RLPLFRDWIKX 851
Db 811 RITGVISWIOQ 821

RESULT 16

CAG30332 PRELIMINARY; PRT; 824 AA.
ID CAG30332
AC CAG30332;
DT 01-JUN-2004 (Tremblrel. 27, Created)
DT 01-JUN-2004 (Tremblrel. 27, Last sequence update)
DT 01-JUN-2004 (Tremblrel. 27, Last annotation update)
DE D11170K4.2 protein.
GN Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Collins J.E., Wright C.L., Edwards C.A., Davis M.P., Grinham J.A.,
 RA Cole C.G., Goward M.E., Aguado B., Mallya M., Mokrab Y., Huckle E.J.,
 RA Beare D.M., Dunham I.,
 RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL:CR456446; CAG30332.1;
 SQ SEQUENCE 824 AA; 91333 MW; A74F186406041F7B CRC64;
 Query Match 23.3%; Score 1092; DB 2; Length 824;
 Best Local Similarity 30.3%; Pred. No. 7,8e-67;
 Matches 276; Conservative 135; Mismatches 330; Indels 170; Gaps 29;
 QY 11 GGPRDFAGLKYKSRHEKNGLEGEVFLPVNNVKKVKGKPGKRVNVLAVLIGLLVLL 70
 DB 11 GGQGDGDDG-----EEAEPEGMFKAEE-----DSKRKARGYLRLVPLFVLL--ALVLA 57
 QY 71 GIGFLVNH-LOYR-DYVYOKYKNGYMRITNENFVDAYENSSTFEVSLASKVOKALKLY 128
 DB 58 SAGVLLWYFLGKYKAEVWSQVYSGSLRVLNHFQDLTRRSSAFRSETATAQKMLKELI 117
 QY 129 SGVPLGPKYHKSAYTAFSEGSVLAAYVSEFIPQHLVEAEARVMAEERVVLPPRARS 188
 DB 118 TSTR-LGTYVSSSVSTGEGPLTCFPPFIQIDBH-----RLMLSPV--V 162
 QY 189 KSPVVTSVVAEPTDSKYVQTODNSCSFGLHARGVELMR-----TTPGPPDSY- 238
 DB 163 QALVEEILSTVNSSAAVPRAE---YEVDPEGLVLEASVKDIAMALNSTLGCYRYSYV 218
 QY 239 -----PAH--ARCOMALRGDADSVSLTFRSPFLASCDRSGSLVTVVNTLSPME 286
 DB 219 GQGVGLRKGDHLASSCLWHLQPKDMLKRL--EWTLACRR--LAMYVAGLE 273
 QY 287 PHALVOL--CGTYPSPVYL-----FHSSQNVLLTLITTERHGFEXTF 331
 DB 274 KRLITSYVCGSRQRPVEVLASGAIMVWKKGLHSYDPPVLV-----QPVV 322
 QY 332 FQLRPMSSCGRLRKAQSTFNSPYPGHYPPNIDCTANIEVNNQHVKKRFYLLPFG 391
 DB 323 FOACEVNLTLDRNLDSQVSLTPYFPSPYSQTHCSWMLTVPSLDYGLALWFDVALRLQ 382
 QY 392 VPACTCEPDYVINGEKYCG-----ERSQFVNTSNKITYTRFHSQDYTOTGLAE 443
 DB 383 KYDLPCTQGWNTINRRLCGRLIOPVAKRIPVAATAG--ITINFSQISLTSPGKRVH 439
 QY 444 YLSYDSDPCPGQFTCTGRCIRKELRCGDWADCTHSDDELNCSGDAGHQTCKNKFCKP 503
 DB 440 YGLYNQSDPCRGELCSV-----NGLCVP 463
 QY 504 LFWVCDVNDGSDNSDEQSCSPAOTFRC-SNGKCLSKSQCGNGKDCGSDGSDAECBPV 562
 DB 464 ---ACDGVKDCPGLIDERNVCRA-TFOCKEDSTCISLPKVCDCQDCLNDSDEQSCD- 518
 QY 563 NVVYCTKATRYCLNGLSKNPECDGDESDSDSDCCGGLRSFRQARVVGCTDAD 622
 DB 519 -GVPCGFTTFQCEBRSCVKKPNQCDGRPDCRDSDEBHCDCGQCG--PSSRIYGAVSS 575
 QY 623 EGEWPMQVSLHALCGNHCASLISPMNLVSAACIYIDRFRYSDDPTQWAFGLHDQS 682
 DB 576 EGEWPMQASLOVRGR-HICGALILADRVITTAHCFOGD---SNASTYLWLVFGLKQWQN 631
 QY 683 QRSAPGVQERLKIISHPFNFDTFYDIALBLEKPAEYSSWVRPICLPDASHVPAG 742
 DB 632 SR-WPEGVSPKYSRLLHPYHEBDSHDYVALLQIDHPVVSAAVRPVCPLARSHFEPG 690
 QY 743 KAIWVTGNGHTQYGTGA-----LILQKGRIVNQTTCENTL 780
 DB 691 LHCWITGALREGLRLDAVALPYGMWNGSESTCCCPISNALQVVDQLPQDLCSEVY 750
 QY 781 PQQTPRMKCVGLSGVDSCGSDSGPLSVSEADGRIFGAGVSWMGDCAGRMKPGVYT 840

DB 751 RYVTPRMLCAGYRKGDACQGDGGLVCKALSGRWFLAGLVSMGLGCGRPVPGVYT 810
 QY 841 RLPLFRMWKE 851
 DB 811 RITGVISWIOQ 821
 RESULT 17
 Q7RTY8 ID Q7RTY8 PRELIMINARY; PRT; 572 AA.
 AC Q7RTY8
 DT 01-MAR-2004 (TREMBLrel. 26, Created)
 DT 01-MAR-2004 (TREMBLrel. 26, Last sequence update)
 DT 01-OCT-2004 (TREMBLrel. 28, Last annotation update)
 DE Type II transmembrane serine protease 7 precursor (hypothetical
 protein FLJ16086).
 GN Name=TPRSS7;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22722134; PubMed=12838346;
 RA Puente X.S., Sanchez L.M., Overall C.M., Lopez-Otin C.,
 RT "Human and mouse proteases: a comparative genomic approach."
 RL Nat. Rev. Genet. 4:544-558(2003).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Nimomiyu K., Magatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
 RA Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,
 RA Katsura N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,
 RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
 RA Kanai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
 RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,
 RA Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
 RA Suzuki Y., Sugano S., Nagahara K., Masuko Y., Nagai K., Isegai T.,
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 CC -1. SIMILARITY: Belongs to peptidase family S1.
 DR EMBL:BN000125; CAD6757.1; -.
 DR EMBL:AK131211; BAD18401.1; -.
 DR MEROPS: S01.072; -.
 DR GO: GO:0016021; C:Integral to membrane; IEA.
 DR GO: GO:0004263; F:chymotrypsin activity; IEA.
 DR GO: GO:0008223; F:peptidase activity; IEA.
 DR GO: GO:0004295; F:trypsin activity; IEA.
 DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro: IPR0010859; CUB.
 DR InterPro: IPR002172; LDL_receptor_A.
 DR InterPro: IPR001254; peptidase_S1.
 DR InterPro: IPR001314; peptidase_S1A.
 DR Pfam: PF00431; CUB; 1.
 DR Pfam: PF00057; Ldl_recep_a; 2.
 DR Pfam: PF00089; Trypsin; 1.
 DR PRINTS: PR00722; CHYMOTRYSIN.
 DR PRINTS: PR00261; LDLRECEPTOR.
 DR PROSITE: PS01180; CUB; 2.
 DR PROSITE: PS01209; LDLRA_1; 1.
 DR PROSITE: PS01068; LDLRA_2; 2.
 DR PROSITE: PS02440; TRYPSIN_DOM; 1.
 DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.
 DR PROSITE: PS00135; TRYPSIN_SER; 1.
 KM HydroLase; Protease; Serine protease; Signal; Transmembrane.
 FT SIGNAL 1 23
 FT CHAIN 24 572 type II transmembrane serine protease 7.
 SQ SEQUENCE 572 AA; 64041 MW; 0458CBEA96EA285 CRC64;
 Query Match 21.4%; Score 1004.5; DB 2; Length 572;
 Best Local Similarity 36.6%; Pred. No. 5.6e-61;
 Matches 228; Conservative 95; Mismatches 223; Indels 77; Gaps 23;
 QY 244 COMALRGDADSVSLTFRSPFLASCDERGSGLVTVVNTLSPMEPHALVOLCGTYPSPVYL 303

```

Db      2 CHEKVAIVGVLIRLSIKSIQI--EADNCVDTSLTIYDSLPIRISILYRIC--EPTKTLM 58
Qy      304 TFFSSQVLLITLTITNTERRRPGEATFEPOLPRMSGSGRLRK-----AOGFPNSPYRGH 359
Db      59 SFTSTNNMLVTFSPHRIIRLSGIRAFEVYIPEQKCNVYLVKQITGPEGKISSPYPSY 118
Qy      360 YPPNIDCTWNIIEVNNQHVKKRKF--YLLEPGVPAGTCPEYVEINGEKYCGE--RSQF 416
Db      119 YPRKCKCTWKEQT--SLSTLIGALKFYNYSTTKSMKG--CEHGMEWEINBHMYCGSYMDHQ 176
Qy      417 VTSNSKKTIVRFPSIDSYDTGTGLAYLSYDSDPQP--GQFTCRTRGCRKEIRCGMA 475
Db      177 IFRVPSPLVHLQLQCSSRLSDKPLLAYGSYNLSQPCVPSPFRCSGILCVPAQRCQGVN 236
Qy      476 DCTHSPSELNCSCDAGHQFTCKNFC---KPLFWVCDSVNDGDNDEGSCSPAQOTFRC 532
Db      237 DCFPESBEL-----FCVSPQP-----ACNTSSFR-- 260
Qy      533 SNGKLSKSQOQCKGDKDGGDSDEASCPKVVVVTCTKHYRCLNGCLSKGNEPCDKED 592
Db      261 QHGPFLI-----CDGFRCDCENGRDQNC--TQSI PCNNRTFCGNDICFRKQNAKCDGTVD 313
Qy      593 CSDSDKDKDCDCLGRSTFRQARVVGTDADGEMPMQVSLHALQCGHICGASLISPMVLV 652
Db      314 CPDSDDEGCTCS--RSSSALHRIIGGDTLEGGMVQVSHFVGSAY--CGASVISREML 371
Qy      653 SAACHYIDDDGFRYSDEPTQMTAFGLHDGORS--APGVDERRLKRIISHPFNDPTFDY 710
Db      372 SAACHF---HGNRISDPTPTAHGMYQGNAKVSP-----VRKLVHETNYISQTFDY 422
Qy      711 DIALLELE--KPAEYSSMVRPILCPDASHVFPACKAIWVGWHT--OYGTGALLIOLKE 767
Db      423 DIALQLQSIAMPETLKQILQICIPPGQVRVSGEKCMTVWGMRHBDKMSLVILQOAE 482
Qy      768 IRVYNQTCENLPLQOTTFRMVCYFUSGVDSCQSGSGLS--SVADGRIFGAYVSW 826
Db      483 VELIDQTLCVSTY--GIITSRMLCGMISGRDACKDGSGLSCRRKSDGKMLITGLVSW 541
Qy      827 GDGCAGRNKPGVYTRPLPLFRDWI 849
Db      542 GHGCGRNPFGVYTRVSNFVPMI 564

RESULT 18
Q8BIK6 PRELIMINARY; PRT: 572 AA.
AC Q8BIK6;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE Mus musculus adult male corpora quadrigemina cDNA, RIKEN full-length
DE enriched library, clone:B230219123 product:weakly similar to BLOOD
DE COAGULATION FACTOR XI.
GN Name=B230219123Rik;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;
RX MEDLINE=99279253; PubMed=10349636;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;
RX MEDLINE=21085660; PubMed=11217851;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;
RA The FANTOM Consortium.
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;
RX MEDLINE=20499374; PubMed=11042159;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;
RX MEDLINE=20530913; PubMed=11076861;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuini T., Taahiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Wataniki M.,
RA Oneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;
RA Aichi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka I., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohnato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazune N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toyota T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Belongs to peptidase family S1.
DR EMBL, AK045663; BAC32448.1; -.
DR HSP, P00760; IEZK.
DR MEROPS, S01.072; -.
DR MGD, MGI:2686594; B230219123Rik.
DR GO, GO:0004263; F:chymotrypsin activity; IEA.
DR GO, GO:0004295; F:trypsin activity; IEA.
DR GO, GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro, IPR000859; CUB.
DR InterPro, IPR002172; LDL_receptor_A.
DR InterPro, IPR001254; Peptidase_S1.
DR InterPro, IPR001314; Peptidase_S1A.
DR InterPro, IPR009003; Pept_Ser_Cys.
DR Pfam, PF00431; CUB. 1.
DR Pfam, PF00057; Ldl_recept_a; 3.
DR Pfam, PF00089; Trypsin. 1.
DR PRINTS, PR00722; CHYMOTRYPSIN.
DR PRINTS, PR00261; LDLRECEPTOR.
DR SMART, SM00042; CUB. 1.
DR SMART, SM00192; LDLa; 3.
DR SMART, SM00020; TYP_Spc. 1.
DR PROSITE, PS01180; CUB. 2.
DR PROSITE, PS01209; LDLA_1; 1.
DR PROSITE, PS00668; LDLA_2; 2.
DR PROSITE, PS00240; TRYPSIN_DOM; 1.
DR PROSITE, PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE, PS00135; TRYPSIN_SER; 1.
DR Hydrolase; Protease; Serine protease.
KW

```

SQ SEQUENCE 572 AA; 63757 MW; 21C0EC523B5F4301 CRC64;
 Query Match 21.4%; Score 1004.5; DB 2; Length 572;
 Best Local Similarity 36.5%; Pred. No. 5.6e-61;
 Matches 226; Conservative 94; Mismatches 229; Indels 71; Gaps 21;

QY 244 COMALRGDADSVLSLTFRSPDLASCDERGSDLTVYNTLSPMEPHALVOLCGTPSPSYNL 303
 DB 2 CHFLVAVIGVYLRLSLKSIQI-EADNCVTDSLTIYDSLPIRSLILYRIC-EPTRTLM 58
 QY 304 TFFSSQVLLITLITNTERRHPGEATFFQLPRMSSCGG-----RLKAOCTFNSPYYPG 358
 DB 59 SFVSTNNMLVTLKSPYRLAGIRAYEVIP-E-QKESTTLVKEINSFEKISSPYPPS 117
 QY 359 HYPNIDCTWNIENVNNOHVVRKFFYLEPGVPAGTCPPDYVEINGEKYCGE--RSQF 416
 DB 118 YPPKCKCTWKFQT-SLSTGLALKFYNYSTTKSKAKCEBGEIWEIHEHMYCGSYMDHET 176
 QY 417 VVTSNSKITYRFRHSDSYTDGFLAEYLSYDSDPCP-GQFTCRGRCIKELRCDGMA 475
 DB 177 IFRVPSPLVHIQLQCSSRLSDKPLAEYGSYNIHQPCVGSFRCSGGLCVPAQRCDGVN 236
 QY 476 DCTDHSDELNCSCDAGHQTCKNKKFC--KPLFWVCDVNDCCGNSDEQSCSPAQTFRC 535
 DB 237 DCFPESDELFCVT-----VXP-----ACNSSFR-QHG 263
 QY 536 KCLSKSQQCKGKDDCGDSDASCPKYNVVTCTKHTYRCLNGCLSKGNEPCDCKEDCD 595
 DB 264 PLV-----CGFRDCBQDQDQNCNTR--SIPCTRTFCGNDICFRKQNAQCDGIVDCPD 316
 QY 596 GSDKDDCGRLSRTRQARVVGTDADGEMPMQVSLHALGQGHICGASLISPMVLSAA 655
 DB 317 GSDEEGCGCS-RSSSPLHRIYGSDDQETPMQVSLHFGVSAV-CGASVISRMLLSAA 374
 QY 656 HCYIDDRGFRYSDEPTQWTAFLGLHDQSQR--APGVERRLKRIISHPFNDFTFDYDIA 713
 DB 375 HCF---HGNRLSDPTPTTAHLGMVYQNAKFI--VRLVHEHYNSQTFDDIA 425
 QY 714 LLELE--KPAEYSSMVRPCLPDASHVFPAGKAIWVGWHT-QYGTGALLILOGEIRV 770
 DB 426 LQLQSIAMPETLKLQIPIPCIPAGOKVRSGCKMVTGWRHEDSKGSFVLQDAEVEL 485
 QY 771 INQTTCNLIPQOITPRPMCVGLSGVSDCGSGGRLS-SVEADGRIFGAGVNSGDG 829
 DB 486 IDQVVCVSTY-GITTSRLCAGVSGSKSDACKDGGSLCRKSDGKMLITGLVSWHG 544
 QY 830 CAGRNKPGVYTRLPLFRDWI 849
 DB 545 CGRNPFGVYTRVSSFVPMI 564

RESULT 19
 BAD18401 PRELIMINARY; PRT; 572 AA.
 AC BAD18401:
 DT 12-MAY-2004 (Tremblrel. 27, Created)
 DT 12-MAY-2004 (Tremblrel. 27, Last sequence update)
 DT 12-MAY-2004 (Tremblrel. 27, Last annotation update)
 DE CDNA FLJ16088 f1s, clone NTZRP7008435, weakly similar to EPITTHIN (EC 3.4.21.-).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Nimomiya K., Nagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
 Rutsuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,
 Katsuta N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,
 RA Otsuki T., Sato K., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
 Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
 RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,
 Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,

RA Suzuki Y., Sugano S., Nagahari K., Masuno Y., Nagai K., Isogai T.;
 RT "NEDO human cDNA sequencing project";
 RL Submitted (MAR2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK31211; BAD18401.1; -
 SQ SEQUENCE 572 AA; 64041 MW; 0458CBEA96BA285 CRC64;

Query Match 21.4%; Score 1004.5; DB 2; Length 572;
 Best Local Similarity 36.6%; Pred. No. 5.6e-61;
 Matches 228; Conservative 95; Mismatches 223; Indels 77; Gaps 23;

QY 244 COMALRGDADSVLSLTFRSPDLASCDERGSDLTVYNTLSPMEPHALVOLCGTPSPSYNL 303
 DB 2 CHFLVAVIGVYLRLSLKSIQI-EADNCVTDSLTIYDSLPIRSLILYRIC-EPTRTLM 58
 QY 304 TFFSSQVLLITLITNTERRHPGEATFFQLPRMSSCGGLK--AQCTFNSPYYPG 359
 DB 59 SFVSTNNMLVTLKSPYRLAGIRAYEVIPQKCENTLVLDITGFEKISSPYPPSY 118
 QY 360 YPPNIDCTWNIENVNNOHVVRKFF-YLLEPGVPAGTCPPDYVEINGEKYCGE--RSQF 416
 DB 119 YPPKCKCTWKFQT-SLSTGLALKFYNYSTTKSKMKG-CEHGEIWEIHEHMYCGSYMDHQT 176
 QY 417 VVTSNSKITYRFRHSDSYTDGFLAEYLSYDSDPCP-GQFTCRGRCIKELRCDGMA 475
 DB 177 IFRVPSPLVHIQLQCSSRLSDKPLAEYGSYNIHQPCVGSFRCSGGLCVPAQRCDGVN 236
 QY 476 DCTDHSDELNCSCDAGHQTCKNKKFC--KPLFWVCDVNDCCGNSDEQSCSPAQTFRC 532
 DB 237 DCFPESDELFCVT-----FCVSPQ-----ACNTSFR- 260
 QY 533 SNKCLSKSQQCKGKDDCGDSDASCPKYNVVTCTKHTYRCLNGCLSKGNEPCDCKED 592
 DB 261 QHGEL-----CGFRDCBQDQDQNCNTR--TQSPCNRRITFCGNDICFRKQNAQCDGIVDCPD 313
 QY 593 GSDKDDCGRLSRTRQARVVGTDADGEMPMQVSLHALGQGHICGASLISPMVLSAA 652
 DB 314 CPDSDEEGCGCS-RSSSALHRIIGGDTLEGMPQVSLHFGVSAV-CGASVISRMLLSAA 371
 QY 653 SAACHYIDDRGFRYSDEPTQWTAFLGLHDQSQR--APGVERRLKRIISHPFNDFTFDY 710
 DB 372 SAACHF---HGNRLSDPTPTTAHLGMVYQNAKFI--VRLVHEHYNSQTFDDIA 422
 QY 711 DLELE--KPAEYSSMVRPCLPDASHVFPAGKAIWVGWHT-QYGTGALLILOGEIRV 767
 DB 423 DIALQSIAMPETLKLQIPIPCIPAGOKVRSGCKMVTGWRHEDSKGSFVLQDAEVEL 482
 QY 768 IRVYNQTTCNLIPQOITPRPMCVGLSGVSDCGSGGRLS-SVEADGRIFGAGVNSG 826
 DB 483 VELIDQTLVSTY-GITTSRLCAGVSGSKRDKDGGSLCRKSDGKMLITGLVSWH 541
 QY 827 GDCGAGRNKPGVYTRLPLFRDWI 849
 DB 542 GHGCGRNPFGVYTRVSNFVPMI 564

RESULT 20
 ENTX_PIG STANDARD; PRT; 1034 AA.
 AC P98074;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Enteropneustidase precursor (EC 3.4.21.9) (Enterokinase).
 GN Name=PRS7; Synonyms=ENTX;
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OX NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
 RA TISSUE=Duodenal mucosa;
 RX MEDLINE=94327548; PubMed=8051081;
 RA Matsushima M., Ichinose M., Yahagi N., Kakei N., Tsukada S., Miki K.,

RA Kurokawa K., Tashiro K., Shickawa K., Shinomiya K., Umeyama H.,
 RA Inoue H., Takahashi T., Takahashi K.,
 RT "Structural characterization of porcine enteropeptidase.";
 RL J. Biol. Chem. 269:19976-19982(1994).
 CC -1- FUNCTION: Responsible for initiating activation of pancreatic
 CC proteolytic proenzymes (trypsin, chymotrypsin and carboxypeptidase
 CC A). It catalyzes the conversion of trypsinogen to trypsin which in
 CC turn activates other proenzymes including chymotrypsinogen,
 CC procarboxypeptidases, and proelastases.
 CC -1- CATALYTIC ACTIVITY: Selective cleavage of 6-Lys-Ile-7 bond in
 CC trypsinogen.
 CC -1- SUBUNIT: Heterotrimer of a catalytic (light) chain, a multidomain
 CC (heavy) chain, and a mini chain.
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
 CC -1- PTM: The chains are derived from a single precursor that is
 CC cleaved by a trypsin-like protease.
 CC -1- PTM: The mini chain may be cleaved by elastase.
 CC -1- SIMILARITY: Belongs to peptidase family S1.
 CC -1- SIMILARITY: Contains 2 CUB domains.
 CC -1- SIMILARITY: Contains 2 LDL-receptor class A domains.
 CC -1- SIMILARITY: Contains 1 MAM domain.
 CC -1- SIMILARITY: Contains 1 SEA domain.
 CC -1- SIMILARITY: Contains 1 SRCR domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
 CC or send an email to license@ebi.ac.uk).
 CC -----
 DR EMBL: D30799; BAA06459.1; -.
 DR HSSP: P98072; IEBK.
 DR MEROPS: S01.156; -.
 DR InterPro: IPR000859; CUB.
 DR InterPro: IPR002172; LDL_receptor_A.
 DR InterPro: IPR000998; MAM.
 DR InterPro: IPR001254; Peptidase_S1.
 DR InterPro: IPR001314; Peptidase_S1A.
 DR InterPro: IPR011163; Pept_S1A_enterop.
 DR InterPro: IPR009003; Pept_Set_Cys.
 DR InterPro: IPR000803; SEA.
 DR InterPro: IPR000082; S1cr_receptor.
 DR InterPro: IPR001190; S1cr_receptor.
 DR Pfam: PF00057; Ldl_recept_a; 2.
 DR Pfam: PF00629; MAM; 1.
 DR Pfam: PF01390; SEA; 1.
 DR Pfam: PF00530; SRCR; 1.
 DR Pfam: PF00089; Trypsin; 1.
 DR PIRSF: PIRSF001138; Enteropeptidase; 1.
 DR PRINTS: PR00722; CHYMOTRYPSIN.
 DR PRINTS: PR00261; LDLRECEPTOR.
 DR PRINTS: PR00020; MAMDOMAIN.
 DR SMART: SM00042; CUB; 2.
 DR SMART: SM00192; LDLa; 2.
 DR SMART: SM00137; MAM; 1.
 DR SMART: SM00200; SEA; 1.
 DR SMART: SM00202; SR; 1.
 DR SMART: SM00020; TYP_SRC; 1.
 DR PROSITE: PS01180; CUB; 2.
 DR PROSITE: PS01209; LDLRA_1; 2.
 DR PROSITE: PS50068; LDLRA_2; 2.
 DR PROSITE: PS00740; MAM_1; 1.
 DR PROSITE: PS50060; MAM_2; 1.
 DR PROSITE: PS50024; SEA; 1.
 DR PROSITE: PS00420; SRCR_1; FALSE_NEG.
 DR PROSITE: PS50287; SRCR_2; 1.
 DR PROSITE: PS50240; TRYPSIN_DOM; 1.
 DR PROSITE: PS00134; TRYPSIN_HIS; 1.
 DR PROSITE: PS00135; TRYPSIN_SER; 1.
 KM Direct protein sequencing; Glycoprotein; Hydrolase; Lipoprotein;
 MYSLcate; Repeat; Serine protease; Signal-anchor; Transmembrane;

KW Zymogen.
 FT CHAIN 52 117
 FT CHAIN 118 799
 FT CHAIN 800 1034
 FT DOMAIN 1 18
 FT TRANSMEM 19 47
 FT DOMAIN 48 1034
 FT DOMAIN 52 169
 FT DOMAIN 197 238
 FT DOMAIN 240 349
 FT DOMAIN 357 519
 FT DOMAIN 539 649
 FT DOMAIN 656 694
 FT DOMAIN 693 786
 FT DOMAIN 800 1034
 FT ACT SITE 840 840
 FT ACT SITE 891 891
 FT ACT SITE 986 986
 FT LIPID 2 2
 FT DISULFID 199 212
 FT DISULFID 206 225
 FT DISULFID 219 236
 FT DISULFID 658 670
 FT DISULFID 665 683
 FT DISULFID 677 692
 FT DISULFID 787 911
 FT DISULFID 825 841
 FT DISULFID 925 992
 FT DISULFID 956 971
 FT DISULFID 982 1010
 FT CARBOHYD 116 116
 FT CARBOHYD 147 147
 FT CARBOHYD 170 170
 FT CARBOHYD 194 194
 FT CARBOHYD 283 283
 FT CARBOHYD 343 343
 FT CARBOHYD 350 350
 FT CARBOHYD 403 403
 FT CARBOHYD 455 455
 FT CARBOHYD 485 485
 FT CARBOHYD 518 518
 FT CARBOHYD 549 549
 FT CARBOHYD 645 645
 FT CARBOHYD 697 697
 FT CARBOHYD 701 701
 FT CARBOHYD 721 721
 FT CARBOHYD 740 740
 FT CARBOHYD 761 761
 FT CARBOHYD 804 804
 FT CARBOHYD 863 863
 FT CARBOHYD 902 902
 FT CARBOHYD 964 964
 SQ SEQUENCE 1034 AA; 114776 MW; 0386C64CF64CC368 CRC64;
 Query Match 15.5%; Score 724; DB 1; Length 1034;
 Best Local Similarity 23.8%; Pred. No. 3; le-11;
 Matches 258; Conservative 135; Mismatches 330; Indels 360; Gaps 41;
 QY 55 WVTLAAILGLILVIGLFLVWHLQYRVRQK--VKNGYWRIT-----NENFVDAYEN 107
 DB 21 FTALFAILWVLCAGLAVSMRLTKGSEKDAALGKSHKRGTKITSGVTYNNPDLQKL-- 78
 QY 108 SNSTEFVSLASKVKDALKLLYSGVPLPGYHKESAVTARSEGSVIAYV-----W-SEPS 160
 DB 79 --SVDFKVLAFDQOMIGETFOSSNLKNEY-KNSRVLOPFENGSVIVIFDLPLPAQWVSDEN 135
 QY 161 IPQHLVEAEARVWAEEVVM----- 180
 DB 136 IKBELIGIIBANKSSGLVAFHIDVNSIDTSELNYSSTSPSTSDKLTSSPATPGNV 195
 QY 181 ----LP--PRARSLKSFV---TSVVAFPDTSKTVQRTQDNCSFGIHHARGVELMRFT 229

```

Db 196 SIECLPSRPAADLAKCIAVDLPDGDGLNCPDSDSDSKICATACD-----GKFLILTES 249
Qy 230 TPFGPDSPYP-----AHARQWALRGDADSVSLTFRSGDLASCDBR-----GSDLV--- 276
Db 250 SGSFDAQYPLSLBASVVCQWIIHVNOGLSIELNFSYFNTYSMVULNIYEGVSSKILRA 309
Qy 277 -----TVVNTLS----- 283
Db 310 SLWLMPGTRIRFISNQVTVTFLEISDENDYIGFNATYAFNSTBLNDEKINCFEDGFC 369
Qy 284 -----PM 285
Db 370 FWIGDLNDNEMERIQGTFPPFGPNPDHFGNASGYISTPGGGRQROVGLSLSLPL 429
Qy 286 EPHALVOLC-----GYPPSYN----- 302
Db 430 EP-TLEPVCLSFWYMYGENVYKLSINISNDQNEKIFQKEGVYGENWNYGVTLNETV 488
Qy 303 ---LTFHSSQVLLITLITNTERHPRF-----EATFF-----QLPRMSSCGG--RLR 345
Db 489 EFKAFAFAFKQFSLDALDDISLTYYGICNVSLYPEPLTVPTSPPELP--TDGCGPELM 546
Qy 346 KAQGFNSPYYPGHYPNIDCTNINIEVNNQHVKVRKFFYLLEPGVPAGTCPRDYEI- 404
Db 547 EPNTFTFSMFPNNYPNAPFCVMNLNMQKNGIQLHPEFPLENIA-----DVEIR 598
Qy 405 NGEK-----YCGERSQFVNTSNKITYRFHSDSYDTGFLAELYS--YDSSDP 453
Db 599 DGEEDDSLALLAVYGPGEVDVFTTNMTVLFTINDALTGKGFANFTTGYHLGIDEP 658
Qy 454 -PGQFTGRTGRCIKELACDGMADCTHSDDELNCSCDAGHPCTCKNKRCKLFWVCDSYN 512
Db 659 KEDNFGQENGEVLLVNLCDGFSHCKGSDAHACV-----RF-----LN 697
Qy 513 DCGNSDEGCGSCPAQTFRC---NGKLSKSOQCNCKDPC---GDGSDASCP----- 560
Db 698 GTANNSG-----LVQFRIGSIWHTACAEWTTQTSIDVQQLGLGIGNSMFPFSG 750
Qy 561 ---KVVVVTCTKTYRCLNGLCLSKNPECDGKED-----CSDGSDKDCDGLSFT 610
Db 751 GPVFKLNTAP-----NGSLILTASEQC--FEDSLILLCQNHKS-----CGKQVA 793
Qy 611 RQA--RVVGGTDADEGEMPVQVSLHALGQGHICGASLISPMWLYSAACHYIDDBGFRSD 668
Db 794 QEVSPKIVGNDSEGAWPVVALYYNGQ--LLCGASLVSRDMLVSAACHYV---RWLE 848
Qy 669 PTQNTAFGLHDORSAPGVQERRLKRIISHPFNDFTFDYDIALLEKPAEYSSMVR 728
Db 849 PSKKKAILGLHMTSNLTSFOIVTRLIDETIVNPHNRRKSDIAMHLEFKVNTDYIQ 908
Qy 729 PICLPDASHVFPACKAIWVTGMHTQYGGTALLQKGEIRVINQTCENLLPQ-QITPR 787
Db 909 PICLPBEMQVPPPRICSIAGMGKVIYCGSPADILQEADVLLNEXKQOQMPXNITEN 968
Qy 788 MNCVGFSLSGVDSQCGSGGSLSSVEADGRIFGAGVSVWGDGCGAKRNPQVYTRLPLFRD 847
Db 969 MMCAGYEBGGIDSCQDGSGLMCLC--NNRWLAGVTSFGYQALPNRPVYAAVPEKTE 1027
Qy 848 WIK 850
Db 1028 WIK 1030

```

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=22784742; Pubmed=12886014;
RA Cal S., Lopez-Otin C.;
RT "Polymerase-I, a human polypotease with the ability to generate
RT independent serine protease domains from a single translation
RT product."
RL Proc. Natl. Acad. Sci. U.S.A. 100:9185-9190(2003).
DR EMBL; AJ488947; CAD35759.1; -.
DR MEROPS; S01.357; -.
DR MEROPS; S01.358; -.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001254; peptidase_S1.
DR Pfam; PF00057; Ldl_recept_a; 1.
DR Pfam; PF00089; Trypsin; 2.
DR SMART; SM00192; LDla; 1.
DR SMART; SM00020; TRYP_SFC; 2.
DR PROSITE; PS50068; LDLRA 2; 1.
DR PROSITE; PS50240; TRYPIN_DOM; 2.
DR PROSITE; PS00134; TRYPIN_HIS; UNKNOWN_2.
DR PROSITE; PS00135; TRYPIN_SER; 2.
KW Hydrolyase; Protease; Serine protease.
SQ SEQUENCE 855 AA; 91352 MW; 8AF2759D9740CF3F CRC64;

Query Match 15.2%; Score 711.5; DB 2; Length 855;
Best Local Similarity 45.7%; Pred. No. 1.8e-40;
Matches 132; Conservative 52; Mismatches 96; Indels 9; Gaps 5;

Qy 567 CTKTYRCLNGLCLSKNPECDGKEDCSDSDKDCDGLRSPTRQ-RVVGCTDADEGE 625
Db 154 CPQNSPFCGNSQCVTKNPNPCDDQEDCSDSDENHCEGLQPAWRMAGRIYGVGEASRGE 213
Qy 626 WPMQVSLHALGQGHICGASLISPMWLYSAACHYIDDBGFRYSPTQNTAFGLHDORS 685
Db 214 FPMQASLIRE-NKEHFCAALINARMLVSAACHFNE---FQDPTKVAAYVATYLSGSE 267
Qy 686 APGVQERRLKRIISHPFNDFTFDYDIALLEKPAEYSSMVRPICLPDASHVFPACKAI 745
Db 268 ASTYR-AQVQIVGHPVYNADFDVAIVLELTSPLFGNHIDPVCLPAATHIFPESKCC 326
Qy 746 WVTGMHTQYGG--TGALLQKGEIRVINQTCENLLPQGITPRMNCVGFSLSGVDSQCGD 804
Db 327 LISGKGLKDFLVKREVLQKATVELLDQALCASLYGHSITDMMVCGIYDQKVDSCQGD 366
Qy 805 SGGFSLSSVEADGRIFGAGVSVWGDGCGAKRNPQVYTRLPLFRDMIKENT 853
Db 387 SGGFLVCEBPSGRFFLAGIVSWGICAEARRPGVYAAVTRLRDIIEAT 435

RESULT 22
Q72411 PRELIMINARY; PRT; 1059 AA.
ID Q72411
AC Q72411;
DT 01-OCT-2003 (Tremblrel. 25, Created)
DT 01-OCT-2003 (Tremblrel. 25, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE Polymerase-1A protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=22784742; Pubmed=12886014;
RA Cal S., Lopez-Otin C.;

```

```

RT "Polypease-1, a human polypease with the ability to generate
RT independent serine protease domains from a single translation
RT product."
RL Proc. Natl. Acad. Sci. U.S.A. 100:9185-9190(2003).
DR EMBL: AJ488946; CAD5758.1; -.
DR MEROPS: S01.357; -.
DR MEROPS: S01.358; -.
DR MEROPS: S01.969; -.
DR GO: GO:0004295; F:trypsin activity; IEA.
DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro: IPR002172; LDL_receptor_A.
DR InterPro: IPR001254; Peptidase_S1.
DR Pfam: Pf00057; Ldl_recept_a; 1.
DR Pfam: Pf00089; Trypsin; 3.
DR SMART: SM00192; Ldlra; 1.
DR SMART: SM00020; Tryp_spc; 3.
DR PROSITE: PS50068; Ldlra_2; 1.
DR PROSITE: PS50240; TRYPSIN_DOM; 3.
DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN; 3.
DR PROSITE: PS00135; TRYPSIN_SER; 2.
KM Hydrolase; Protease; Serine protease.
SQ SEQUENCE 1059 AA; 114020 MW; 17D27A2D9F2A264 CRC64;

Query Match 15.2%; Score 711.5; DB 2; Length 1059;
Best Local Similarity 45.7%; Pred. No. 2.3e-40;
Matches 132; Conservative 52; Mismatches 96; Indels 9; Gaps 5;

QY CTKHYRCLNGLCLSKNPECDCGKEDSCGDEKCCGRLSPFROA-RVVGTDADGE 625
Db 154 CPGNSPFCGNSQCTTKNPECDDESDSDSEHCEGLOPAMRMAGRVGMEASPG 213
QY 626 WPMQVSLHALGQGHICGASLISPMVLVAHAYCIDRGRFYSDPTQWTAFLGHDGORS 685
Db 214 FPMQASLRE-NKEHFCAALINAMVLVAHCFNE-----FQDPTKWAVYAGATYLSGE 267
QY 686 APGQGERLKRILHPFNPTPYDIALLEPRATSSVNRPICLPDASHVPACKAI 745
Db 268 ASTVR-AQVAVQIVGHPYMDAPDAVAVLELTPFGRHIIQVCLPAATHIPPSKCC 326
QY 746 WVTSGHQTQYGG-IGALILIGKEIRVINTQTCENLPPQITPRMVCVGLSGVDSQGD 804
Db 327 LISMGVLKEDFLVKEPVQLKATYELDQALCASLHSHSLTDRAVCGAGYLDGKVDSCGD 386
QY 805 SGGPLSSVEADGRIFGAGVSWGDGCAGRNKGYYTRLPFRMIKENT 853
Db 387 SGGPLVCEPSSGRFFLAGIVSWGICAEARRPGYAVTRLRDWILBAT 435

RESULT 23
ENTR BOVIN STANDARD; PRT; 1035 AA.
ID ENTR BOVIN STANDARD; PRT; 1035 AA.
AC P98072;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE Enteropetidase precursor (EC 3.4.21.9) (Enterokinase).
GN Name=PR587; Synonyms=ENTK;
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovinae; Bos.
OC NCBI_TaxID=9913;
RX NCBI_TaxID=9913;
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Duodenum;
RA Kltamco Y., Yuan X., Wu Q., McCourt D.W., Sadler J.B.;
RT "Enterokinase, the initiator of intestinal digestion, is a mosaic
RT protease composed of a distinctive assortment of domains."
RL Proc. Natl. Acad. Sci. U.S.A. 91:7588-7592(1994).
RN [2]
RP SEQUENCE OF 801-1035 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=94043122; PubMed=8226855;

```

```

RA Lavallie E.R., Rehmetulla A., Racie L.A., DiPasio E.A., Ferenz C.,
RA Grant K.L., Light A., McCoy J.M.;
RT "Cloning and functional expression of a cDNA encoding the catalytic
RT subunit of bovine enterokinase."
RL J. Biol. Chem. 268:23311-23317(1993).
RN [3]
RP SEQUENCE OF 801-827.
RC TISSUE=Intestine;
RA MEDLINE=92189715; PubMed=1799406;
RX Light A., Janska H.;
RT "The amino-terminal sequence of the catalytic subunit of bovine
RT enterokinase."
RL J. Protein Chem. 10:475-480(1991).
CC -1- FUNCTION: Responsible for initiating activation of pancreatic
CC proteolytic proenzymes (trypsin, chymotrypsin and carboxypeptidase
CC A). It catalyzes the conversion of trypsinogen to trypsin which in
CC turn activates other proenzymes including chymotrypsinogen,
CC procarboxypeptidases, and proelastases.
CC -1- CATALYTIC ACTIVITY: Selective cleavage of 6-Lys-|-Ile-7 bond in
CC trypsinogen.
CC -1- SUBUNIT: Heterodimer of a catalytic (light) chain and a
CC multidomain (heavy) chain linked by a disulfide bond.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=P98072-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=P98072-2; Sequence=VSP_005386;
CC -1- TISSUE SPECIFICITY: Intestinal brush border.
CC -1- PFM: The chains are derived from a single precursor that is
CC cleaved by a trypsin-like protease.
CC -1- SIMILARITY: Belongs to peptidase family S1.
CC -1- SIMILARITY: Contains 2 CUB domains.
CC -1- SIMILARITY: Contains 2 LDL-receptor class A domains.
CC -1- SIMILARITY: Contains 1 SEA domain.
CC -1- SIMILARITY: Contains 1 SEA domain.
CC -1- SIMILARITY: Contains 1 SRCR domain.
CC -1- SIMILARITY: Contains 1 SRCR domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U09859; AAB40026.1; -.
DR EMBL: L19663; AAI16035.1; -.
DR PIR: A43090; A43090.
DR PDB: 1EKX; X-ray; A=788-800, B=801-1035.
DR MEROPS: S01.156; -.
DR InterPro: IPR000859; CUB.
DR InterPro: IPR002172; LDL_receptor_A.
DR InterPro: IPR000998; MAM.
DR InterPro: IPR001254; Peptidase_S1.
DR InterPro: IPR001314; Peptidase_S1A.
DR InterPro: IPR001163; Pept_S1A_enterop.
DR InterPro: IPR009003; Pept_Set_Cys.
DR InterPro: IPR000082; SEA.
DR InterPro: IPR001190; Srcr_receptor.
DR Pfam: Pf00431; CUB; 2.
DR Pfam: Pf00057; Ldl_recept_a; 2.
DR Pfam: Pf00629; MAM; 1.
DR Pfam: Pf01390; SEA; 1.
DR Pfam: Pf00530; SRCR; 1.
DR Pfam: Pf00089; Trypsin; 1.
DR PIRSF: PIRSF001138; Enteropetidase; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR PRINTS: PR00261; LDLRECEPTOR.
DR PROSITE: PS01180; CUB; 2.
DR PROSITE: PS01209; LDLRA_1; 2.
DR PROSITE: PS50068; LDLRA_2; 2.

```


DR	PROSITE; PS00740; MAM 1; 1.	
DR	PROSITE; PS50050; MAM_2; 1.	
DR	PROSITE; PS50024; SEA; 1.	
DR	PROSITE; PS00420; SRCR 1; FALSE_NEG.	
DR	PROSITE; PS50287; SRCR 2; 1.	
DR	PROSITE; PS50240; TRYPsin DOM; 1.	
DR	PROSITE; PS00134; TRYPsin HIS; 1.	
DR	PROSITE; PS00135; TRYPsin_SER; 1.	
KM	3d-structure; Alternative splicing; Direct protein sequencing;	
KM	glycoprotein; Hydrolyase; Lipoprotein; Myristate; Repeat;	
KM	Serine protease; Signal-anchor; Transmembrane; Zymogen.	
FT	CHAIN 1 800	
FT	CHAIN 801 1035	
FT	DOMAIN 1 18	
FT	TRANSMEM 19 47	
FT	DOMAIN 48 1035	
FT	ACT_SITE 154 169	
FT	DOMAIN 197 238	
FT	DOMAIN 240 330	
FT	DOMAIN 358 520	
FT	DOMAIN 540 650	
FT	DOMAIN 657 695	
FT	DOMAIN 694 787	
FT	DOMAIN 801 1035	
FT	ACT_SITE 841 841	
FT	ACT_SITE 892 892	
FT	ACT_SITE 987 987	
FT	LIPID 2 2	
FT	DISULFID 199 212	
FT	DISULFID 206 225	
FT	DISULFID 219 226	
FT	DISULFID 659 671	
FT	DISULFID 666 684	
FT	DISULFID 678 693	
FT	DISULFID 788 912	
FT	DISULFID 826 842	
FT	DISULFID 926 932	
FT	DISULFID 957 972	
FT	DISULFID 983 1011	
FT	CARBOHYD 116 116	
FT	CARBOHYD 147 147	
FT	CARBOHYD 170 170	
FT	CARBOHYD 194 194	
FT	CARBOHYD 233 233	
FT	CARBOHYD 263 263	
FT	CARBOHYD 264 264	
FT	CARBOHYD 404 404	
FT	CARBOHYD 456 456	
FT	CARBOHYD 486 486	
FT	CARBOHYD 519 519	
FT	CARBOHYD 550 550	
FT	CARBOHYD 646 646	
FT	CARBOHYD 698 698	
FT	CARBOHYD 722 722	
FT	CARBOHYD 741 741	
FT	CARBOHYD 762 762	
FT	CARBOHYD 864 864	
FT	CARBOHYD 903 903	
FT	CARBOHYD 965 965	
FT	VARSPLIC 166 192	
FT	CONFLICT 808 808	
FT	STRAND 802 802	
FT	STRAND 805 806	
FT	TURN 809 810	
FT	TURN 813 814	
FT	STRAND 815 820	
FT	TURN 821 822	
FT	STRAND 823 830	
FT	STRAND 835 838	
FT	HELIX 840 843	
FT	TURN 844 845	
R -> Y (in Ref. 3).		

FT	HELIX	850	852
FT	STRAND	853	857
FT	STRAND	861	861
FT	TURN	862	863
FT	TURN	868	869
FT	STRAND	871	880
FT	TURN	882	883
FT	STRAND	885	885
FT	TURN	886	889
FT	STRAND	890	890
FT	STRAND	894	898
FT	STRAND	912	912
FT	TURN	916	917
FT	TURN	922	923
FT	STRAND	925	930
FT	TURN	933	933
FT	STRAND	936	937
FT	STRAND	940	940
FT	STRAND	943	943
FT	STRAND	945	951
FT	HELIx	954	960
FT	TURN	962	963
FT	TURN	968	969
FT	STRAND	970	973
FT	TURN	976	977
FT	STRAND	981	981
FT	TURN	984	985
FT	STRAND	987	988
FT	TURN	990	995
FT	TURN	996	997
FT	STRAND	998	1007

Query Match	15.2%; Score 709.5; DB 1; Length 1035;
Beet Local Similarity	23.1%; Pred. No. 3,1e-40;
Matches 251; Conservative 150; Mismatches 320; Indels 365; Gaps 44	

OY	55	WVYLAAVLIGLLVLGIGFLVMHLQYRDV--RVQKAKNGMKRTITNENFVDAYENSNET	111
Db	21	FAVFILIIIVLACAGLIAVSWLISOGSVDAAFGKSHERGLTKIISGATYNPHLDKLSV	80
OY	112	BFVSLASKVKALKLVSGVPFLGPYHKESAVTAPSEGSVIAYV-----W-SEPSIPOH	164
Db	81	DFKVLAFTIQMIDDIFFOSSNLKNEY-KNSRYLOENGSIIVIFDLPLPQMWSDKNVKEE	139
OY	165	LVEAEARVMAEBERYVM-----LPPARSLKSFWVTGSV-----V	197
Db	140	LIQGIEANKSSQLVTFHIDLMSIDITASLENSTISPATTSEK-LTTISIPLATPGNVS	197
OY	198	AEPPTDSKTVRTODNSCFGLHARGVELM-----LPPTPGF	233
Db	198	ECPSPDSSLR-----CADALCKIAIDLFCDEGLNCPDGSDDEDNKTCATACDCGRFLLTGS	249
OY	234	PDS-----PYPAH---ARCQWALRGDADSVLSLTFERSPD-----	264
Db	250	SGSFPEALHYPKPSNNTSAVCRWITIKVNOGLSTQLNFDFNYTYADVLANIYEGMGSKILIR	309
OY	265	-----LASCDERGSDLV---TVYNTLSPMIEPHALVOL-----	293
Db	310	ASLMSNNPCGILIRIFSNQVTATFLIOSDE--SYDIGFKYLYTAFNSEKELANNYEKINCNFPED	367
OY	294	--C-----GTTPSPYNLTFRHS-----	307
Db	368	GFCFPIQIOLDNDNEWERTQGSTFPSTGPTFDHTGENESGFISTPTGCGRRERRVGLLT	427
OY	308	-----SONVLLITLITNERKHHPGEAATFFC-----	333
Db	428	LPLDPTPEOACLSEFYMYNGENVYKLISNISDOF--NMEXTIFQKEGNYGNMNMYGYVT	484
OY	334	-----LPRMSSCGG	342
Db	485	LNETVTEFKVSRYGFGNQILSDIALDDISLTGYICANVSYPREPTLYVPTPPLELP--TDCCG	542
OY	343	--RLUKAQGTFSPPYPGHYPPENIDCTWNIEBVNPQNQHVRKREFYLLBPGVDPACTCPXD	400

```

Db      543 PHDMENNTFTSINFPNSYPNQAFCIMWNAOKGKXIOHFOFDEJENIA-----D 594
Qy      401 YVEI-NEKXCGERSOV-----YTSNKNKTIVFPHSQSTYDGLAEYLS-- 446
Db      595 VVEIRDSB---GDSLSFLAYTTGPGPVNDVFTSTNMTVLFITIMNAKQGFANFTTGY 651
Qy      447 -YDSDDC-PGQFTCRTRGRCIRKELRCDGMADCTDHSDELNC-----SCDAGH--OPTC 496
Db      652 GLGIPBCKEDNFOCKDGECPILVNLCDGPHCKDGDDEAHCVRLFNGTDDSSGLVQFRI 711
Qy      497 KAKFCRKLFWVCDSDVNDGNSDQSCSAPQTRCGNKGKLSQCGNCKDQDSDSDE 556
Db      712 QS-----IMHV-----ACAEN-----MTTQISDDVC-----QLLG-----LGTNSS 743
Qy      557 ASC-----PRVNVVTCFKATYRCIANGLCISKNPECDKEDCSGSDKDCD---CGLR 607
Db      744 VPTFTSGGPGVYNLNTAP-----NCSLITPEQC-----LEDLSLILQCNKSCGCK 791
Qy      608 SFTROA--RVVGTDADEGEWPMQVSLHALGQGHICGASLISPMNLVSAACHYIDRGR 665
Db      792 LVTOEVSFKIYVSGDSREGAMPWVALYPDDQ--QVCGASLVSRLMVSAAHCYVG-----R 846
Qy      666 YSDPTQMTAFLGLHDQGRAPVQERLRKRIISHPPNDFTPYDIALLEKPAEYSS 725
Db      847 NMESKMKAVLGLHMASNLTSPOLETRLIDQIVINPHYNRKRKNNDIAMHLEKRVYTD 906
Qy      726 MVRPICLPDASHVFPAGKAIWVTGMHTQYGTGALILQKGEIVINQTTCEMLPQ-QI 784
Db      907 YIQPICPENQVFPFRIGSISAGMALIYQGSINADVLOQEDVLSNEKQCOQMPENYI 966
Qy      785 TPRMVCGLISGVVSCQSGSGEPLSSVEADGRIFGAGVSWDGCAGRRKPGVYTLPL 844
Db      967 TENNVCAGYBAGVDSQCGSGGPIMCQE--NNRWILAGVTSFGYQCALPVRPGYAVPR 1025
Qy      845 FRDMIK 850
Db      1026 FTEWIQ 1031

RESULT 24
CORI_HUMAN STANDARD; PRT; 1042 AA.
AC Q9Y5Q5; Q9UHY2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 05-JUN-2004 (Rel. 44, Last annotation update)
DE Atrial natriuretic peptide-converting enzyme (EC 3.4.21.-) (pro-ANP-
DE converting enzyme) (Corin) (Heart specific serine proteinase ATC2).
GN Name=CORIN; Synonyms=CRN;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=99262646; PubMed=10329693;
RA Yan W., Sheng N., Seto M., Morser J., Wu Q.;
RT "Corin, a mosaic transmembrane serine protease encoded by a novel cDNA
RT from human heart."
RL J. Biol. Chem. 274:14926-14935 (1999).
RN [2]
RP SEQUENCE OF 734-1040 FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=20534769; PubMed=11082206;
RA Hooper J.D., Scorman A.L., Clarke B.E., Normyle J.F., Antalio T.M.;
RT "Localization of the mosaic transmembrane serine protease corin to
RT heart myocytes."
RL Eur. J. Biochem. 267:6931-6937 (2000).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=20359740; PubMed=10880574; DOI=10.1073/pnas.150149097;

```

```

RA Yan W., Wu F., Morser J., Wu Q.;
RT "Corin, a transmembrane cardiac serine protease, acts as a pro-atrial
RT natriuretic peptide-converting enzyme."
RL Proc. Natl. Acad. Sci. U.S.A. 97:8525-8529 (2000).
CC -1- FUNCTION: Converts pro-ANP to ANP. Cleaves pro-ANP specifically
CC between Arg-123 and Ser-124.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein.
CC -1- TISSUE SPECIFICITY: Highly expressed in heart. Expressed in heart
CC myocytes.
CC -1- SIMILARITY: Belongs to peptidase family S1.
CC -1- SIMILARITY: Contains 2 LDL-receptor (F2) domains.
CC -1- SIMILARITY: Contains 7 LDL-receptor class A domains.
CC -1- SIMILARITY: Contains 1 SRCR domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; A013845; AAD31850.1; -.
DR EMBL; A0113248; AAF21966.1; -.
DR HSP; P01130; IAUU.
DR MEROPS; S01.019; -.
DR Genew; HGNC:19012; CORIN.
DR MIM; 605236; -.
DR GO; GO:0005887; C:integral to plasma membrane; TAS.
DR GO; GO:0004252; F:serine-type endopeptidase activity; TAS.
DR GO; GO:0006629; P:lipid metabolism; TAS.
DR GO; GO:0006629; P:lipid metabolism; TAS.
DR GO; GO:0006508; P:proteolysis and peptidolysis; TAS.
DR GO; GO:0008217; P:regulation of blood pressure; TAS.
DR InterPro; IPR000024; F2 domain.
DR InterPro; IPR001372; LDL_receptor_A.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR InterPro; IPR009003; Pept_Ser_Cys.
DR InterPro; IPR001190; Srcr_receptor.
DR Pfam; PF01392; Fz_2.
DR Pfam; PR00057; Idl_recept_a; 6.
DR Pfam; PR00089; Trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00261; LDLRECEPTOR.
DR PROSITE; PS50038; Fz_2.
DR PROSITE; PS01209; LDLRA_1; 6.
DR PROSITE; PS01068; LDLRA_2; 7.
DR PROSITE; PS00420; SRCR_1; FALSE_NEG.
DR PROSITE; PS50287; SRCR_2; FALSE_NEG.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; FALSE_NEG.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Glycoprotein; Hydrolase; Repeat; Serine protease; Signal-anchor;
KW Transmembrane.
KX DOMAIN 1 45
FT TRANSMEM 46 66
FT SIGNAL 67 1042
FT DOMAIN 134 259
FT DOMAIN 268 304
FT DOMAIN 305 340
FT DOMAIN 341 377
FT DOMAIN 378 415
FT DOMAIN 450 573
FT DOMAIN 579 614
FT DOMAIN 615 653
FT DOMAIN 654 690
FT DOMAIN 690 786
FT DOMAIN 802 1042
FT ACT_SITE 843 843
FT ACT_SITE 892 892
FT ACT_SITE 985 985
FT ACT_SITE 985 985

```

```
FT DISULFID 790 912 By similarity.
FT DISULFID 828 844 By similarity.
FT DISULFID 955 970 By similarity.
FT DISULFID 981 1010 By similarity.
FT CARBOHYD 80 80 N-linked (G1CNAC . . .) (Potential).
FT CARBOHYD 104 104 N-linked (G1CNAC . . .) (Potential).
FT CARBOHYD 135 135 N-linked (G1CNAC . . .) (Potential).
FT CARBOHYD 141 141 N-linked (G1CNAC . . .) (Potential).
FT CARBOHYD 231 231 N-linked (G1CNAC . . .) (Potential).
FT CARBOHYD 245 245 N-linked (G1CNAC . . .) (Potential).
FT CARBOHYD 251 251 N-linked (G1CNAC . . .) (Potential).
FT CARBOHYD 305 305 N-linked (G1CNAC . . .) (Potential).
FT CARBOHYD 320 320 N-linked (G1CNAC . . .) (Potential).
FT CARBOHYD 376 376 N-linked (G1CNAC . . .) (Potential).
FT CARBOHYD 413 413 N-linked (G1CNAC . . .) (Potential).
FT CARBOHYD 446 446 N-linked (G1CNAC . . .) (Potential).
FT CARBOHYD 451 451 N-linked (G1CNAC . . .) (Potential).
FT CARBOHYD 469 469 N-linked (G1CNAC . . .) (Potential).
FT CARBOHYD 567 567 N-linked (G1CNAC . . .) (Potential).
FT CARBOHYD 651 651 N-linked (G1CNAC . . .) (Potential).
FT CARBOHYD 697 697 N-linked (G1CNAC . . .) (Potential).
FT CARBOHYD 761 761 N-linked (G1CNAC . . .) (Potential).
FT MUTAGEN 1022 1022 N-linked (G1CNAC . . .) (Potential).
FT CONFLICT 854 854 S->A: Loss of activity.
FT CONFLICT 876 876 W->R (in Ref. 2).
SQ SEQUENCE 1042 AA; 116564 MW; 7705398B8607AD2 CRC64;

Query Match 14.8%; Score 695; DB 1; Length 1042;
Best Local Similarity 28.2%; Pred. No. 3,1e-39;
Matches 210; Conservative 97; Mismatches 265; Indels 172; Gaps 32;

Qy 202 DSKTVQRTQDNQSCFGLHARVEMLRFTTGPFPDPPYPAHARQOMALRGDAD----- 253
Db 365 DHDCVDSDEVNCS--CHSQGLVECR-----NSQCIPTSTQCC-----GDDECKDGDDE 412
Qy 254 --SVLSTLFRSFD-----LASCDERGSDLTVVNTL--SPMEPHALVOLCGTVPSPS 300
Db 413 NCSYIQTSQCGDRCLYNPLDSCG--GSPLCDPNNSLNCSQCEPTTL-ELCMNLP-- 467
Qy 301 YNLTF-----HSSQNVLLITLINTERRHGFATFQPLRMSSCGGRLLKAQGTENSP 354
Db 468 YNYSYNYNFGHRQKASIS-----WESSLFPAVQVNCYKYLMPFECTI--- 513
Qy 355 YYPGHYPNIDCTNNIEVN---NQYKVRKFFYLIEPVPAGTCPKDYVEINGEKYC 410
Db 514 ----LVPKCDVNTGERIPPCRALCEHSKERCESVL-----GIVGLQWPEDE----- 554
Qy 411 GERSQFVNTSNKRTVRFHSDQSYTDGFLAEVLSYDSSDPC--PGQFTCTGRCIKREL 469
Db 555 TDCSQFPEBNSDQTCI-----MDEVY-----ECSFSHFKCRSGQCVLASR 597
Qy 470 RCDGMADCTHSDSLNCSCDAGHOFTC--KNFKCRPLFVNCDSVNDGSDQSCS--CPA 527
Db 598 RCDGQACDDSDSDENCGCKERDLMEGCSNKKCLKHTYICGFPDCCPYMDKKNCSFGQD 657
Qy 528 QTFRCNSGKCLSKSQCCNGKDDCGSDDEASCPKVN----- 564
Db 658 DELECANHACVSRDLWCDGADSCDSDEMDCVLSTIVNSSFELMVHARATEHHVCAQD 717
Qy 565 -----VTC-----TKHT--YKCLNGCLSK-----GNPBDG 589
Db 718 WQELISQACQKQWGLGEBPSYTKLIQEDKEPRMLTLHSNMSLNGTTLHELLVNGQSCS 777
Qy 590 KEDCSDSDEKDCDCCGLRSFTR--QARVVGTDADEGEWPMQVSLHALGQGHICASLISP 648
Db 778 RSKSISLCTKQ--DQGRPARPMKRLIGRTSRGPRPWCGLSQSBSGHICCVILAK 835
Qy 649 NMLVSAHCTYDDRGFRYSPTQWTAFLGLHDSQSORAPGVQERLKRILSHPFENFTF 708
Db 836 KMWLTVAHCF--EGR-----ENAAVWKVVGINNLDHPSV--FMQTRFVKIILHPPYSRAVV 889
Qy 709 DYDIALLEKRPAYSSWVRPICLPDASHVFPAGKALVNTGWHGTQVGGTALLQKGEI 768
```

```
Db 890 DYDLSYELBEDISEYGVAVPVCLPNEQWLEPPTYCYITGWH--NGNKMPFLQSGEV 947
Qy 769 RVINQTTCEMLLPQO--ITPRMVCYGLSGVDSQGDGSGPGLSVEDGRIFGAGVSWG 827
Db 948 RIISLEHCQGVFDMKRTITTMICAGYESGTVDSCMGDSGGLVCEKPGHWTLLGILTSWG 1007
Qy 828 DGCAGR--NKPGVYRLLPLFRDWIK 850
Db 1008 SVCPKVLGPGVSNVSVFVEMIK 1031

RESULT 25
AC Q80YN4 PRELIMINARY; PRT; 1111 AA.
ID Q80YN4;
DC 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Corin.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley;
RA Langenickel T., Pagel I., Buttgeriet J., Tenner K., Lindner M.,
RL Willebrock R., Dietz R., Bader M.;
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Belongs to peptidase family 11.
DR EMBL; AY251285; AA086772.1; -.
DR HSBP; Q07954; ICR8.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0004044; F:scavenger receptor activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0007275; P:development; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000024; Fz_domain.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR InterPro; IPR009003; Pept_Set_Cys.
DR InterPro; IPR001190; Strc_Receptor.
DR Pfam; PF01392; Fz_2.
DR Pfam; PF00057; Ldl_recept_a; 6.
DR Pfam; PF00089; Trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00261; LDLRECEPTOR.
DR SMART; SM00063; FRI; 2.
DR SMART; SM00192; LDLa; 7.
DR SMART; SM00202; SR; 1.
DR SMART; SM00202; TRYP_SPC; 1.
DR PROSITE; PS50038; Fz_2.
DR PROSITE; PS01209; LDLRA_1; 6.
DR PROSITE; PS50068; LDLRA_2; 7.
DR PROSITE; PS50287; SRCR_2; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolyase; Protease; Serine protease.
SQ SEQUENCE 1111 AA; 122635 MW; 3BA2706CBE81157F CRC64;

Query Match 14.8%; Score 694.5; DB 2; Length 1111;
Best Local Similarity 28.7%; Pred. No. 3,7e-39;
Matches 210; Conservative 80; Mismatches 259; Indels 183; Gaps 31;

Qy 205 TVQPTQNGSCFGLHARVEMLRFTTGPFPDPPYPAHARQOMALRGDADSVLSTFRSFD 264
Db 462 TFCQDGDGDECKDSDENCS-----DRPFP-----CPGDRGGLDS----- 497
Qy 265 LASGDE--RGSDDLTVVNTL--SPMEPHALVOLCGTVPSPSYNLTFP----- 306
```

```

Db 498 --SCVESCAAGSLCSDSDSLNSCHCEPITL-ELCMNLP--YNLTHYPNLYGHRTQKEAS 552
Qy 307 -SSQNVLLITLITTERRHPFEATPFOLPRMSSCGRLKRAQGTFSYPYGHYP-NI 364
Db 553 ISMSALFPALVQNCVKYLMFPACTILVFKCDVNTQ-----RVPPCRLL 597
Qy 365 DCTWNIIEVPNNQHVKVRPKFPYLLPEGVPACTCKDYVEINGEKXCGERSQFVNTSNK 424
Db 598 LC-----EHSKRCESVL-----GIVGLQMPED-----TDCSQFPQSSDNQ 634
Qy 425 IYRFHSDQYTDYGFLEAYLVSDDSPC-RGQTCRTGRCKRLKELACDGNADCTHSD 483
Db 635 TCL-----LPNEDVEBCSPSHKCRSGRCVLSRRCDGQADDDSD 677
Qy 484 LNSCDAHQFTCK-NKFEKPLFWCVSDVNDGNSBOGCS-CPAOTFRCSNGKCLSKS 541
Db 678 ENGCKCRDLMEGCLNNQCLKHTLICGFPCCSMBEKXCSFQDDELBECANECYPRD 737
Qy 542 QQCNKGKDDCGDGPBASCPRKVVNTCTKH-----TYRCLNG-----L 578
Db 738 LMCQGWTDSCSDSDMGC-----VTLKNGSSSFLTVHRSARDHYACADGMOETLSOLA 792
Qy 579 C--LSKNPE-----CDKE-----DSSGSDSK-----D 601
Db 793 CRQGLSEPVTBLVQCGEQQMLRLHSSWENLNGSTLQELVHRSCPSGSEISLCTK 852
Qy 602 CDCGLRSFTT-QARVVGTDADDEGMPWOVSLHALGQGHICGSLISPNMLVSAHGYID 660
Db 853 QDCGRRAAAMNKRILIGRTSRPRMWPQCSLQSBGSHICGCLAKKWLTVAHCF-E 911
Qy 661 DRGRYDPTQWTAFLGLHDSQSRAPGVORRLKRIISHPFENDFTFYDIALLEKRP 720
Db 912 GR----EDAVYWKVYFGINNDHPSG-FMQTRFVKTLLHPRYSRAVVDYDVSVELSD 966
Qy 721 AEYSWTRPCLPDASVFPAGKAIWYGMGHTGYGALILQKGRVYNQTCENL 780
Db 967 INETSYRVPCLPSPREFLEBDTYCYITGMCH--MKNMPPKLOEGEERILPLEQCSYF 1024
Qy 781 PQQ-ITPRMVCVGLSGVDSCQDSCGAPLSSVEADRIFAGVSWGDCAGR-NKPGV 838
Db 1025 DMKITTRMTCAGVSGTVDSCMGDSGPLVCERPGQWTLFGLTSGSVCFSLVLRPGV 1084
Qy 839 YTRLPLFRDWIK 850
Db 1085 YSNVSYFVDWIE 1096

```

```

RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RA The FANTOM Consortium;
RT "Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=2049374; PubMed=11042159;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RN Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RL prepare full-length cDNA libraries for rapid discovery of new genes.";
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=20530913; PubMed=11076861;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RN Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Smi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsunoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Matsubiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RL sequencing pipeline with 384 multicapillary sequencer.";
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RN Fukuda S., Furuta M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashizaki K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RN Hori F., Imclati K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RN Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RN Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shingawa A., Shiraki T., Sogabe Y., Tagami M.,
RN Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Belongs to peptidase family 11.
DR EMBL; AK038356; BAC29973.1; -.
DR MGD; MGI:1197523; P1887.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0016301; F:kinase activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0005044; F:scavenger receptor activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000858; CUB.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR000998; MAM.
DR InterPro; IPR01254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR InterPro; IPR009003; Pept_Ser_Cys.
DR InterPro; IPR001190; SrcF_receptor.
DR Pfam; PF00431; CUB; 2.
DR Pfam; PF00057; Ldl_recept_a; 1.
DR Pfam; PF00629; MAM; 1.
DR Pfam; PF00530; SRCR; 1.
DR Pfam; PF00899; Trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.

```

```

DR PRINTS: PRO0020; MAMDOMAIN.
DR SMART: SMO0042; CUB; 1.
DR SMART: SMO0192; LDM; 1.
DR SMART: SMO0137; MAM; 1.
DR SMART: SMO0202; TRP; 1.
DR SMART: SMO0020; TRP_SPC; 1.
DR PROSITE: PS01180; CUB; 1.
DR PROSITE: PS01209; LDM; 1.
DR PROSITE: PS50068; LDM; 1.
DR PROSITE: PS50060; MAM; 1.
DR PROSITE: PS50287; SRCR; 2.
DR PROSITE: PS50240; TRYPsin; 1.
DR PROSITE: PS00134; TRYPsin; 1.
DR PROSITE: PS00135; TRYPsin; 1.
DR Glycoprotein; Hydroxylase; Kinase; Protease; Serine protease.
KW NON TER
SQ SEQUENCE 777 AA; 87314 MW; A18E2F4ECF0ED3A8 CRC64;

Query Match 14.6%; Score 684; DB 2; Length 777;
Best Local Similarity 29.3%; Pred. No. 1.3e-38;
Matches 186; Conservative 93; Mismatches 220; Indels 136; Gaps 22;

Qy 258 LTFRSFLASCDERGSGLVTVYNTLSPMEPHALVOLCTYPPSYNLTFHSSQNVLLITLI 317
Db 231 VVFYAFNRRCSTALDIDISTLNGICSGSPYEPPLVTPPP----- 272
Qy 318 TNRERHPRFEATFQLEPRMSSCGG--RLRAQGTENSPYFGHYPPNIDCTNIEVPPNN 375
Db 273 -----ELP--TDCGPFELMEPNSTFSSPNPDYKYPQASCIWMLNQRG 315
Qy 376 QHVVRKFPFLLEPRVAGCPDYEI--NGE-----KYGESQFVVTNSKIT 426
Db 316 KNIOHLOEFLN-----INDVVEVRDGEFDSLLAVYGPGRVQDLFTFTMT 367
Qy 427 VRFSDSYDTDTGLAEYLIS---YDSDPG--PGQFTCTGRCIRKELKCDMADCTHSD 482
Db 368 VIPTNMTETRRKGRKANFTSGYIGIPEPCDDDFQCDGNCICILGNLCDSYHPHCRGSD 427
Qy 483 ELNC-----SCDAGHQTCKNFKCPFLWCDVNDGSDNSDEQSCPAOTFRCSN 534
Db 428 EASCVRFLNGTRSNNGLVQFNHS-----TWHI-----ACAEN-----WTQISN 467
Qy 535 GKCLSKSQOCCNGKDDCGDGSDEACSP-----KYNVYTCIKHTYRCINGLCLSKGN 584
Db 468 EVC-----HLLGSGANSSMPISSTGGGPFVRN-----QAPNGSLILTPS 508
Qy 585 PECGKE---DCSDGSEKDCDGLSFTROA--RVVGTDADEGEWPMQVSLHALGQG 638
Db 509 LQCSQDSDILLCNHS-----CGEKVTKVSPKIVGSDAQAGAMPVVALYHNDRS 562
Qy 639 H--ICGASLISPMVLVSAAHCYIDRGRFYSPTQWTAFLGLHDGQSRAPGVQERRLK 695
Db 563 TDRLLCGLASIVSDMLVSAHCVRR---RMLDTRMTAVLGLHQSMLTSPQVVRVRVD 618
Qy 696 RIISHPRFNDTFRYDIALLELEKPAEYSSMVRICLPDASHVPRAGALVNTGHTQY 755
Db 619 QIVINPHYDRRRKVNDAIMMLERKVNVTYDIPCLPREEQVITPGKTSIAWGWDKI 678
Qy 756 GGTGALLIKGEIIVINQTCENTLPI-QITPRMVCVFLSGVDSCGDGGPLSSVEA 814
Db 679 NGSTVDVPLKADVLSNEKCOQLPEVNTESMICAGYEKGIDSCDGGGFLMCE- 737
Qy 815 DGRIFGAGVSWGDCAGRNKPGYVYTRLPFRMI 849
Db 738 NNRWFLVGVTSFGVOCALFNHGVYVRSQIEMI 772

```

RESULT 27
 ENTX_HUMAN STANDARD; PRT; 1019 AA.
 AC P98073;
 DT 01-FEB-1996 (Rel. 33, Created)

```

DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Enteropeptidase precursor (EC 3.4.21.9) (Enterokinase).
GN Name=PRSS7; Synonyms=ENTK;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euteria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Duodenum;
RX MEDLINE=95234679; Pubmed=7718557;
RA Kitamoto Y., Velle R.A., Doms-Keller H., Sadler J.E.;
RT "cDNA sequence and chromosomal localization of human enterokinase, the
RT proteolytic activator of trypsinogen."
RL Biochemistry 34:4562-4568(1995).
RN [2]
RP SEQUENCE FROM N.A. AND DISEASE.
RX MEDLINE=21606074; Pubmed=11719902;
RA Holzinger A., Maier E.M., Buck C., Mayerhofer P.U., Kappler M.,
RA Haworth J.C., Moroz S.P., Hadorn H.-B., Sadler J.E., Roscher A.A.;
RT "Mutations in the proenteropeptidase gene are the molecular cause of
RT congenital enteropeptidase deficiency."
RL Am. J. Hum. Genet. 70:20-25(2002).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=20289799; Pubmed=10830953;
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA Pak H.-S., Toyoda A., Ishii K., Tokoi Y., Choi D.-K., Groner Y.,
RA Soeda E., Onki M., Takagi T., Sakaki Y., Tauden S., Blechschmidt K.,
RA Polley A., Menzel U., Delbar J., Kumpf K., Lehmann R., Patterson D.,
RA Reichwald K., Rump A., Schillabel M., Schudy A., Zimmermann W.,
RA Rosenthal A., Kudon J., Shibuya K., Kawasaki K., Asakawa S.,
RA Shintani A., Sasaki T., Negamine K., Mitsuyama S., Antonakakis S.E.,
RA Minoshima S., Shimizu N., Nordstiek G., Hornischer K., Brandt P.,
RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Blocker H.,
RA Ransner J., Beck A., Klages S., Hennig S., Rieseemann L., Dagand E.,
RA Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA Lehrach H., Reinhardt R., Yaspo M.-L.,
RL "The DNA sequence of human chromosome 21."
RL Nature 405:311-319(2000).
RN [4]
RP SEQUENCE OF 749-1019 FROM N.A.
RX TISSUE=Duodenum;
RX MEDLINE=94329561; Pubmed=8052624;
RA Kitamoto Y., Yuan X., Wu Q., McCourt D.W., Sadler J.E.;
RT "Enterokinase, the initiator of intestinal digestion, is a mosaic
RT protease composed of a distinctive assortment of domains."
RL Proc. Natl. Acad. Sci. U.S.A. 91:7588-7592(1994).
CC -1- FUNCTION: Responsible for initiating activation of pancreatic
CC proteolytic proenzymes (trypsin, chymotrypsin and carboxypeptidase
CC A). It catalyzes the conversion of trypsinogen to trypsin which in
CC turn activates other proenzymes including chymotrypsinogen,
CC procarboxypeptidases, and proelastases.
CC -1- CATALYTIC ACTIVITY: Selective cleavage of 6-Lys-|-Ile-7 bond in
CC trypsinogen.
CC -1- SUBUNIT: Heterodimer of a catalytic (light) chain and a
CC multimeric (heavy) chain linked by a disulfide bond.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein (probable).
CC -1- TISSUE SPECIFICITY: Intestinal brush border.
CC -1- PFM: The chains are derived from a single precursor that is
CC cleaved by a trypsin-like protease.
CC -1- DISEASE: Defects in PRSS7 are a cause of enterokinase deficiency
CC [MIM:226200]; a life-threatening intestinal malabsorption disorder
CC characterized by diarrhea and failure to thrive.
CC -1- SIMILARITY: Belongs to peptidase family S1.
CC -1- SIMILARITY: Contains 2 CUB domains.
CC -1- SIMILARITY: Contains 1 LDL-receptor class A domains.
CC -1- SIMILARITY: Contains 1 SEA domain.
CC -1- SIMILARITY: Contains 1 SRCR domain.
CC -1- SIMILARITY: Contains 1 SRCR domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
CC or send an email to license@ebi.ac.uk).

CC -----
DR EMBL; U09860; AAC50138.1; -
DR EMBL; Y19124; CAB65555.1; -
DR EMBL; Y19125; CAB65555.1; JOINED.
DR EMBL; Y19126; CAB65555.1; JOINED.
DR EMBL; Y19127; CAB65555.1; JOINED.
DR EMBL; Y19128; CAB65555.1; JOINED.
DR EMBL; Y19129; CAB65555.1; JOINED.
DR EMBL; Y19130; CAB65555.1; JOINED.
DR EMBL; Y19131; CAB65555.1; JOINED.
DR EMBL; Y19132; CAB65555.1; JOINED.
DR EMBL; Y19133; CAB65555.1; JOINED.
DR EMBL; Y19134; CAB65555.1; JOINED.
DR EMBL; Y19135; CAB65555.1; JOINED.
DR EMBL; Y19136; CAB65555.1; JOINED.
DR EMBL; Y19137; CAB65555.1; JOINED.
DR EMBL; Y19138; CAB65555.1; JOINED.
DR EMBL; Y19139; CAB65555.1; JOINED.
DR EMBL; Y19140; CAB65555.1; JOINED.
DR EMBL; Y19141; CAB65555.1; JOINED.
DR EMBL; Y19142; CAB65555.1; JOINED.
DR EMBL; Y19143; CAB65555.1; JOINED.
DR EMBL; AL163217; CAB90389.1; -
DR EMBL; AL163217; CAB90389.1; -
DR PIR; A56318; A56318.
DR HSSP; P98072; 1EKB.
DR MEROPS; S01.156; -
DR Genes; HGNC:9490; PRSS7.
DR MIM; 226200; -
DR GO; GO:0005903; C:brush border; TAS.
DR InterPro; IPR000857; CUB.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR000998; MAM.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR011163; Pept_S1A_enterop.
DR InterPro; IPR009003; Pept_Set_Cys.
DR InterPro; IPR000082; SEA.
DR InterPro; IPR001190; Srcr_receptor.
DR Pfam; PF000431; CUB; 2.
DR Pfam; PF00057; LDL_recept_a; 2.
DR Pfam; PF00629; MAM; 1.
DR Pfam; PF01390; SEA; 1.
DR Pfam; PF00530; SRCR; 1.
DR Pfam; PF00089; Trypsin; 1.
DR PIRSF; PIRSF001138; Enteropetidase; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00261; LDLRECEPTOR.
DR PRINTS; PR00020; MAMDOMAIN.
DR PROSITE; PS01180; CUB; 2.
DR PROSITE; PS01209; LDLRA_1; 2.
DR PROSITE; PS50068; LDLRA_2; 2.
DR PROSITE; PS00740; MAM_1; 1.
DR PROSITE; PS50060; MAM_2; 1.
DR PROSITE; PS50024; SEA; 1.
DR PROSITE; PS00420; SRCR_1; FALSE_NEG.
DR PROSITE; PS50287; SRCR_2; 1.
DR PROSITE; PS50240; TRYPsin_DOM; 1.
DR PROSITE; PS00134; TRYPsin_HIS; 1.
DR PROSITE; PS00135; TRYPsin_SER; 1.
KW Glycoprotein; Hydroxylase; Lipoprotein; Myristate; Repeat;
KW Serine protease; Signal-anchor; Transmembrane; Zymogen.
FT CHAIN 1 784 Enteropetidase non-catalytic chain.
FT DOMAIN 1 1019 Enteropetidase catalytic chain.
FT TRANSMEM 19 47 Signal-anchor for type II membrane

FT DOMAIN 48 1019 protein (Potential).
FT DOMAIN 52 169 Extracellular (Potential).
FT DOMAIN 182 223 SEA.
FT DOMAIN 225 334 LDL-receptor class A 1.
FT DOMAIN 342 504 CUB 1.
FT DOMAIN 524 634 MAM.
FT DOMAIN 641 679 CUB 2.
FT DOMAIN 678 771 LDL-receptor class A 2.
FT DOMAIN 785 1019 SRCR.
FT ACT_SITE 825 825 Serine protease.
FT ACT_SITE 876 876 Charge relay system (By similarity).
FT ACT_SITE 971 971 Charge relay system (By similarity).
FT LIPID 2 N-myristoyl glycine (Potential).
FT DISULFID 184 197 By similarity.
FT DISULFID 191 210 By similarity.
FT DISULFID 204 221 By similarity.
FT DISULFID 643 655 By similarity.
FT DISULFID 650 668 By similarity.
FT DISULFID 662 677 By similarity.
FT DISULFID 772 896 Interchain (By similarity).
FT DISULFID 810 826 By similarity.
FT DISULFID 910 977 By similarity.
FT DISULFID 941 956 By similarity.
FT DISULFID 967 995 By similarity.
FT CARBOHYD 116 116 N-linked (GlcNAc...)
FT CARBOHYD 147 147 N-linked (GlcNAc...)
FT CARBOHYD 179 179 N-linked (GlcNAc...)
FT CARBOHYD 328 328 N-linked (GlcNAc...)
FT CARBOHYD 335 335 N-linked (GlcNAc...)
FT CARBOHYD 388 388 N-linked (GlcNAc...)
FT CARBOHYD 440 440 N-linked (GlcNAc...)
FT CARBOHYD 470 470 N-linked (GlcNAc...)
FT CARBOHYD 503 503 N-linked (GlcNAc...)
FT CARBOHYD 534 534 N-linked (GlcNAc...)
FT CARBOHYD 630 630 N-linked (GlcNAc...)
FT CARBOHYD 682 682 N-linked (GlcNAc...)
FT CARBOHYD 706 706 N-linked (GlcNAc...)
FT CARBOHYD 725 725 N-linked (GlcNAc...)
FT CARBOHYD 848 848 N-linked (GlcNAc...)
FT CARBOHYD 887 887 N-linked (GlcNAc...)
FT CARBOHYD 909 909 N-linked (GlcNAc...)
FT CARBOHYD 949 949 N-linked (GlcNAc...)
FT CONFLICT 134 134 Q -> E (in Ref. 3).
FT CONFLICT 732 732 S -> P (in Ref. 3).
FT CONFLICT 771 771 SOQLDSLRLDLCNHNKS -> RRNAKEIDALSPILLIA (in Ref. 3).
SQ SEQUENCE 1019 AA, 112923 MM, BCAA245F6DA563 CRC64;
Query Match 14.5%; Score 679.5; DB 1; Length 1019;
Best local similarity 32.5%; Pred. No. 3, 6e-38;
Matches 180; Conservative 80; Mismatches 202; Indels 91; Gaps 20;
QY 333 QLPRMSSCGG--RLRAAGCFHNSPYRPGHYPPRIDCTWNEVNNQHKVRFEPYLLER 330
Db 519 ELP--IDCGGPEFLMEPPNTTFSSTNPNSTPNLAFCWMIINAQKGIQLHPEFDLEN- 575
QY 391 GVPAGTCPEKDYVEI-NGEK-----YCGERSQFVYTSNSKITYRFHSDQSYDTGFL 441
Db 576 -----INDVVEIRGEENDSLAVYTPGKVVXVFSTNNMTYLLITNDVLARGGFK 628
QY 442 AEYL---YDSSDPCFG-OFTCTGRCIRKEILRCGMAADCTHSDLEINCSDAQHOF--- 494
Db 629 ANFTTGYHLGIPRCADHFGQCKNGECVPLVNLCDHLCEDESDADCV-----RFPNG 683
QY 495 TCKN---KRCKLFLVVCDSVNDGCGNSDQGSCEAQGFRGNGGCLSGSQCKNKDCC 550
Db 684 TNNNGLVRRRIGSTI---HTACAE-----WTTQISNDVC-----QLL 719
QY 551 GDSDASACPKVAVVCTKHTYRCLNGLCISKNGPECDG-----KEDSCDGS-----D 598
Db 720 GLDSGNSKRIPT-----DGGPFRVKTATPADGHLITLPQGCQLQDSLIRLQCN 768

```

Oy 599 EKDCDCGLRFTROARVVGSTDADEGEWPMQVSLHALQGGHICGASLISPMNLVSAACY 658
Dy 769 HKSCGCKLAADITPKIVGSSNAKEGAMPVWVGLY-YGRLICGASLVSMTLVSAACV 827
Oy 659 IDDRGFYSYDPTQWTAFLGLHDQSORAPGVQERLKRILSHPFNFDPDIALDELE 718
Dy 828 YG----NRLBSKMTALIGLHMKSNLTSPOVPLRIDEIVNPHYNNRRKNDIAMHLE 883
Oy 719 KPAEYSSMWRIPICPDASHVPAGKAIWVTWGHITQVGGTALLQKEIVINQTCEN 778
Dy 884 PKVAVTYIPIPICPPEBNQVPPPRNCSIAQMGTVVVGSTVANTLQEAADVPLSNRCCQ 943
Oy 779 LLPO-OTPRMAGVFLSGVDSQCGSGPLSSVEADGRIFGAGVSWMGCGAGRNKPG 837
Dy 944 QMEPYNTENNICAGYERGGIDSCQDGGGLMCOE-NNRWFLAGVTSFGIKCALPNRPG 1002
Oy 838 VYTRLPLEFRDWIK 850
Dy 1003 VYARVSRFTWIQ 1015

RESULT 28
ENTK_MOUSE STANDARD; PRT; 1069 AA.
AC P97435;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Enteropeptidase (EC 3.4.21.9) (Enterokinase).
GN Name=Prs87; Synonyms=Entk;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Duodenum;
RX MEDLINE=98147142; PubMed=9486188;
RA Yuan X., Zheng X., Lu D., Rubin D.C., Pung C.Y.M., Sadler J.E.;
RT "Structure of murine enterokinase (enteropeptidase) and expression in
RT Am. J. Physiol. 274:G342-G349(1998)".
RL
CC -1- FUNCTION: Responsible for initiating activation of pancreatic
CC proteolytic proenzymes (trypsin, chymotrypsin and cathepsin)
CC A). It catalyzes the conversion of trypsinogen to trypsin which in
CC turn activates other proenzymes including chymotrypsinogen,
CC procarboxypeptidases, and proelastases (By similarity).
CC -1- CATALYTIC ACTIVITY: Selective cleavage of 6-Lys-Ile-7 bond in
CC trypsinogen.
CC -1- SUBUNIT: Heterodimer of a catalytic (light) chain and a
CC multimeric (heavy) chain linked by a disulfide bond (By
CC similarity).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
CC -1- PTM: The chains are derived from a single precursor that is
CC cleaved by a trypsin-like protease (By similarity).
CC -1- SIMILARITY: Belongs to peptidase family S1.
CC -1- SIMILARITY: Contains 2 CUB domains.
CC -1- SIMILARITY: Contains 2 LDL-receptor class A domains.
CC -1- SIMILARITY: Contains 1 MAM domain.
CC -1- SIMILARITY: Contains 1 SEA domain.
CC -1- SIMILARITY: Contains 1 SRCR domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U73378; AAB37317.1; -.
DR HSSP; P98072; IEKB.
DR MEROPS; S01.156; -.

```

```

DR MD; MG1.1197523; Prs87.
DR InterPro; IPR000859; CUB.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR000998; MAM.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR InterPro; IPR011163; Pept_S1A_enterop.
DR InterPro; IPR009003; Pept_Ser_Cys.
DR InterPro; IPR000082; SEA.
DR InterPro; IPR011190; Srcr_receptor.
DR Pfam; PF00431; CUB; 2.
DR Pfam; PF00057; Ldl_recept_a; 2.
DR Pfam; PF00629; MAM; 1.
DR Pfam; PF01390; SEA; 1.
DR Pfam; PF00530; SRCR; 1.
DR Pfam; PF00089; Trypsin; 1.
DR PRINTS; PRS0722; CHYMOTRYPSIN.
DR PRINTS; PRS0261; LDLRECEPTOR.
DR PRINTS; PRS0020; MAMDOMAIN.
DR SMART; SM00042; CUB; 2.
DR SMART; SM00192; LDLA; 2.
DR SMART; SM00137; MAM; 1.
DR SMART; SM00200; SEA; 1.
DR SMART; SM00202; SR; 1.
DR SMART; SM00020; TRY_SPC; 1.
DR PROSITE; PS01180; CUB; 2.
DR PROSITE; PS01209; LDLRA_1; 2.
DR PROSITE; PS50068; LDLRA_2; 2.
DR PROSITE; PS00740; MAM_1; 1.
DR PROSITE; PS50060; MAM_2; 1.
DR PROSITE; PS50024; SEA; 1.
DR PROSITE; PS00420; SRCR_1; FALSE_NEG.
DR PROSITE; PS50287; SRCR_2; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Glycoprotein; Hydrolase; Repeat; Serine protease; Signal-anchor;
KW Transmembrane; Zymogen.
FT CHAIN 1 829
FT CHAIN 830 1069
FT DOMAIN 1 18
FT TRANSMEM 19 47
FT DOMAIN 48 1069
FT DOMAIN 52 169
FT DOMAIN 227 268
FT DOMAIN 270 379
FT DOMAIN 387 549
FT DOMAIN 569 679
FT DOMAIN 686 724
FT DOMAIN 723 816
FT DOMAIN 830 1069
FT ACT_SITE 874 874
FT ACT_SITE 925 925
FT ACT_SITE 1021 1021
FT DISULFID 229 242
FT DISULFID 236 255
FT DISULFID 249 266
FT DISULFID 688 700
FT DISULFID 695 713
FT DISULFID 707 722
FT DISULFID 817 945
FT DISULFID 859 875
FT DISULFID 959 1007
FT DISULFID 991 1006
FT DISULFID 1017 1045
FT CARBOHYD 147 147
FT CARBOHYD 197 197
FT CARBOHYD 212 212
FT CARBOHYD 373 373
FT CARBOHYD 380 380
FT CARBOHYD 433 433

```

SO	SEQUENCE	1069	AA:	118735	MM:	E6254925463743C3D	CRC64:
FT	CARBOHYD	515	515	N-linked	(G1CNAC . .)	(Potential).	
FT	CARBOHYD	579	579	N-linked	(G1CNAC . .)	(Potential).	
FT	CARBOHYD	675	675	N-linked	(G1CNAC . .)	(Potential).	
FT	CARBOHYD	727	727	N-linked	(G1CNAC . .)	(Potential).	
FT	CARBOHYD	751	751	N-linked	(G1CNAC . .)	(Potential).	
FT	CARBOHYD	770	770	N-linked	(G1CNAC . .)	(Potential).	
FT	CARBOHYD	791	791	N-linked	(G1CNAC . .)	(Potential).	
FT	CARBOHYD	897	897	N-linked	(G1CNAC . .)	(Potential).	
FT	CARBOHYD	936	936	N-linked	(G1CNAC . .)	(Potential).	
FT	CARBOHYD	999	999	N-linked	(G1CNAC . .)	(Potential).	
SO	SEQUENCE	1069	AA:	118735	MM:	E6254925463743C3D	CRC64:

Query Match	14.4%;	Score 673.5;	DB 1;	Length 1069;
Best Local Similarity	29.2%;	Pred. No. 1e-37;		
Matches 186;	Conservative 93;	Mismatches 220;	Indels 137;	Gaps 23;

QY	258	LTFSFSLASCDERKSDLVTVYNNLSPMEPALVOLCTYPPSNLTFHSSQNLITLYI	311
Dd	522	VVFNAFNRGSGTTLALDIDLSTLNGICSGSPPEPLVTPPP-----	563
QY	318	TNTRRHGFATFEPOLPERMSSCGG--RLRAAQFENSRYRGHPRVIDCTWNI-EVENN	375
Dd	564	-----ELP--TDCGPEFLMEPNSTFSPNFDDKTPNOASCTWNLINAROG	606
QY	376	QHVAKRFEPYLLBPGVPAGTCPAOYAI--NGE-----KYGERSQFVNTSNNKIT	426
Dd	607	KNIDLHFEPLDEN-----INDVVEVRDGERDLSLLAVLYGPGPVKOLLFTSTNMT	658
QY	427	VRFSDDGVTDTGFLAELYLS---YDSDPG-C-PGQFTCRGRCIRKELRCDGMADCTHSD	482
Dd	659	VIFTTNMETRKGRKANFTSGYIIGCIPPCODDEFQCKDNCICPLGNLCBDSYPRCRDSD	718
QY	483	ELNC-----SCDAGHOFTCKNFKFCLFWVCDSDVNDGDNDSQSCSPADTFRCSN	534
Dd	719	EASCVRFNGTRSNNGVLQVFNHS-----TWHI-----ACAEH-----WTTQISN	758
QY	535	GKCLSKSQOQNGKDDCGDSEACSP-----KVNVTCTKHFTYRCNLGSLSGN	584
Dd	759	EVC-----HLLGIGANSNSMPSISTGGPFVRAN-----QANGSLILTPS	799
QY	585	PECGKE---DCSDGDEKDCDGLASFTRQA--RVVGCTDADBGEMFPOYSLHALGQG	638
Dd	800	LQCSQDSLILLQCNHKS-----GGEKKVTQCSBKIVGSGSDAAGAAMPWVALYHNRDS	853
QY	639	H---ICGASLISPMWLVSAAHCYCIDDRFXSYDPTQWAFGLHDSQSRASPAVOERLTK	695
Dd	854	TDRLLCGASIVSDMWLSAAHCYR---RLNDPTRWAVLGLHMQSULTSPQVYRVRVD	909
QY	696	RIISHPFNDTFDYDILLLELEKPAEYSSMWPRPLCLPDASHVRPAGALVWTGHTQY	755
Dd	910	QIVINPHDRRRKVNDDIAMMHLERKVNVTYDIQPLCPRENOQIFPGTCSIAAGXDKI	969
QY	756	-GGGALILAGEIRVINVNOTTCENILPQ-QITPRMVCGFISGGVDSQSGSGEPFLSVE	813
Dd	970	NAGSTVDVLKADAVPLLSNEKCOQOOLPEYNTITESMIGAYEEGGIDSCQDSGGFLMQE	1029
QY	814	ADGBIFGAGVYSKMGDCAGRKRPQYTRRLPLFRDMI	849
Dd	1030	-NNRWFVLGVTSFEGVOCALPNHGPYGVAVSOFIEHT	1064

RESULT	29
CORI_MOUSE	
ID_CORI_MOUSE	STANDARD; PRT; 1113 AA.
AC	Q9Z3I9;
DT	16-OCT-2001 (Rel. 40, Created)
DT	16-OCT-2001 (Rel. 40, Last sequence update)
DT	05-JUL-2004 (Rel. 44, Last annotation update)
DE	Artrial natriuretic peptide-converting enzyme (BC 3.4.21.-) (pro-ANP-converting enzyme) (corin) (Low density lipoprotein receptor related protein 4).
GN	Name=Corin; Synonyms=Crm, Lrp4;
OS	Mus musculus (Mouse).

OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX	NCBI_TaxID=10090;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=98429596; PubMed=9756624;
RA	Tonita Y., Kim D.-H., Magoori K., Fujino T., Yamamoto T.T.;
RT	"A novel low-density lipoprotein receptor-related protein with type II membrane protein-like structure is abundant in heart.";
RL	J. Biochem. 124:784-789(1998).
CC	-1- FUNCTION: Converts pro-ANP to ANP. Cleaves pro-ANP specifically between Arg-122 and Ser-123 (By similarity).
CC	-1- SUBCELLULAR LOCATION: Type II membrane protein.
CC	-1- TISSUE SPECIFICITY: Highly expressed in heart.
CC	-1- SIMILARITY: Belongs to peptidase family S1.
CC	-1- SIMILARITY: Contains 2 frizzled (FZ) domains.
CC	-1- SIMILARITY: Contains 7 LDL-receptor class A domains.
CC	-1- SIMILARITY: Contains 1 SRCR domain.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.ebi.ac.uk/announcements/)
CC	or send an email to license@ebi.ac.uk .
CC	-----
DR	EMBL; AB013874; BAA34371.1; -.
DR	PIR; JEO315; JEO315.
DR	HSP; P01130; JALJ.
DR	MEROPS; S01.019; -.
DR	MGD; MGI:1349451; Cortin.
DR	InterPro; IPR000024; Fz_domain.
DR	InterPro; IPR002172; LDL_receptor_A.
DR	InterPro; IPR001254; Peptidase_S1.
DR	InterPro; IPR001314; Peptidase_S1A.
DR	InterPro; IPR009003; Pept_Ser_Cys.
DR	InterPro; IPR001190; Srcr_receptor.
DR	Pfam; PF01392; Fz_2.
DR	Pfam; PF00057; Ldl_recept_a; 6.
DR	Pfam; PF00089; Trypsin; 1.
DR	PRINTS; PR00722; CHYMOTRYPSIN.
DR	PRINTS; PR00261; LDLRECEPTOR.
DR	SMART; SMART0063; FRI_2.
DR	SMART; SMART192; LDLa; 7.
DR	SMART; SMART0202; SR; 1.
DR	SMART; SMART0020; Tryp_Spc; 1.
DR	PROSITE; PS50038; FZ_2.
DR	PROSITE; PS01209; LDLRA_1; 6.
DR	PROSITE; PS50068; LDLRA_2; 7.
DR	PROSITE; PS00420; SRCR_1; FALSE_NEG.
DR	PROSITE; PS50287; SRCR_2; 1.
DR	PROSITE; PS50240; TRYPSIN_DOM_1.
DR	PROSITE; PS00134; TRYPSIN_HIS; FALSE_NEG.
DR	PROSITE; PS00135; TRYPSIN_SER_1.
KW	Glycoprotein; Hydrolase; Repeat; Serine protease; Signal-anchor; Transmembrane.
KW	DOMAIN 1 112 Cytosolic (Potential),
FT	TRANSMEM 113 133 Signal-anchor for type II membrane protein (Potential).
FT	EXTRACELL 1 Extracellular (Potential).
FT	DOMAIN 134 1113 LDL-receptor class A 1.
FT	DOMAIN 201 327 LDL-receptor class A 2.
FT	DOMAIN 336 372 LDL-receptor class A 3.
FT	DOMAIN 373 408 LDL-receptor class A 4.
FT	DOMAIN 409 445 FZ 2.
FT	DOMAIN 446 483 LDL-receptor class A 5.
FT	DOMAIN 518 641 LDL-receptor class A 6.
FT	DOMAIN 647 682 LDL-receptor class A 7.
FT	DOMAIN 683 721 LDL-receptor class A 8.
FT	DOMAIN 722 757 LDL-receptor class A 9.
FT	DOMAIN 758 853 SRCR.
FT	DOMAIN 869 1113 Serine protease.
FT	ACT_SITE 910 910 Charge relay system (By similarity).


```

FT ACT SITE 959 959 Charge relay system (By similarity).
FT ACT_SITE 1052 1052 Charge relay system (By similarity).
FT DISULFID 857 979 By similarity.
FT DISULFID 895 911 By similarity.
FT DISULFID 1022 1037 By similarity.
FT DISULFID 1048 1077 By similarity.
FT CARBOHYD 147 147 N-linked (G1CNAc . .) (Potential).
FT CARBOHYD 202 202 N-linked (G1CNAc . .) (Potential).
FT CARBOHYD 208 208 N-linked (G1CNAc . .) (Potential).
FT CARBOHYD 298 298 N-linked (G1CNAc . .) (Potential).
FT CARBOHYD 317 317 N-linked (G1CNAc . .) (Potential).
FT CARBOHYD 373 373 N-linked (G1CNAc . .) (Potential).
FT CARBOHYD 411 411 N-linked (G1CNAc . .) (Potential).
FT CARBOHYD 444 444 N-linked (G1CNAc . .) (Potential).
FT CARBOHYD 481 481 N-linked (G1CNAc . .) (Potential).
FT CARBOHYD 519 519 N-linked (G1CNAc . .) (Potential).
FT CARBOHYD 537 537 N-linked (G1CNAc . .) (Potential).
FT CARBOHYD 635 635 N-linked (G1CNAc . .) (Potential).
FT CARBOHYD 719 719 N-linked (G1CNAc . .) (Potential).
FT CARBOHYD 765 765 N-linked (G1CNAc . .) (Potential).
FT CARBOHYD 828 828 N-linked (G1CNAc . .) (Potential).
FT CARBOHYD 970 970 N-linked (G1CNAc . .) (Potential).
FT CARBOHYD 1089 1089 N-linked (G1CNAc . .) (Potential).
SQ SEQUENCE 1113 AA; 122984 MW; B845B2C5F20DD8EC CRC64;

Query March 14.1%; Score 661.5; DB 1; Length 1113;
Best Local Similarity 32.7%; Pred. No. 7.1e-37;
Matches 158; Conservative 67; Mismatches 187; Indels 71; Gaps 16;

Qy 428 RFSHDSQYTDGTGFLAEYLSDSPDPCQFTCRGCRKELRCGMADCTDHSDELNCS 487
Db 627 QFPESSDNDQCLLP---NEDVECSPEHFKCRSGRCVLSRRCDGADCDDEBENCG 683
Qy 488 CDAGHQFTCK-NKFCCKPLFWVCDVNDGDSDEQCS-CPAQTFRCNGKCLSKSQCN 545
Db 684 CKEBALMECPFNKCOLKHTLLICDGFPPCDPSMDEKNCSFCODNELECANHECVPRDLCD 743
Qy 546 GKDDCGSGSDASCPKVN-----VYTCNHTYCLNG-----LC----- 579
Db 744 GWVDSDDSDDEMGCVTLISKNGSSLLTVHKSAAEHVCAQWRETTLSQLACKOMGLGEP 803
Qy 580 -LSKGNPECDKE-----DCSDGSDKED-----C--DCGLSFT 610
Db 804 SVTLILGEGEGQMLRLVPMNENLNGSLQELVYRHSCEPSREISLCSQDGRRAA 863
Qy 611 R-QARVVGTDADGEEMFQVSLNALGGHICGSLISPNLVSAHCYIDDRGRYSDP 669
Db 864 RMNKRILIGRTSRGRPMWCSSLQSEPSGHICGCVLAKKWLIVAHCF-EGR---EDA 918
Qy 670 TQWTAFLGHLHQDSQARSAGVERLRKRIISHPFNDFPDVDIALLEKAEVSSWVR 729
Db 919 DVWKEVFGINLLDHPG-FMQTRFVKTLILHPRYSRAVVDYDISVELSDINETSYP 977
Qy 730 ICLPDASHVFAKAIWVTGNGHTQYGGTGAALLQKGRIRVYNOTTCENLLPQO-ITPRM 788
Db 978 VCLSPSEYLEPDTYCYITGNGH--MKNMPEKIQEGSVRLIIPLEQCSYIDMKITTKM 1005
Qy 789 MCVFPLSGVDSCGDSGGLPLSSVEADGRIFGAGVNSGDCAGR-NKPGVYTLPLFRD 847
Db 1036 ICAGYEGSTVDSQMGDSGGLVCEPBGQWTLFGLTSGSVCFKSLGPGVYSNVSYVG 1095
Qy 848 MIR 850
Db 1096 WIE 1098

```

```

DE Embryonic serine protease-2.
GN Name-Xesp-2;
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N. A.
EX MEDLINE=20363741; PubMed=10903452;
RA Yamada K., Takahatake T., Takeshima K.;
RT "Isolation and characterization of three novel serine protease genes
from Xenopus laevis."
RL Gene 252:209-216 (2000).
CC -!- SIMILARITY: Belongs to peptidase family S1.
EMBL: AB038497; BAB08217.1; -.
DR HSSP; P00760; IEZX.
DR MEROPS; S01.049; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0005044; F:scavenger receptor activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR002172; LDL receptor A.
DR InterPro; IPR001254; Peptidase S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR InterPro; IPR009003; Pept_Ser_Cys.
DR Pfam; PF00057; Ldl_recept_a; 4.
DR Pfam; PF00089; Trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00261; LDLRECEPTOR.
DR SMART; SM00192; LDLra; 8.
DR SMART; SM00020; Tryp_spc; 1.
DR PROSITE; PS01209; LDLRA 1; 8.
DR PROSITE; PS00068; LDLRA 2; 2.
DR PROSITE; PS02040; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR Hydrolase; Protease; Serine protease.
SQ SEQUENCE 767 AA; 86001 MW; E0566A38796DE96E CRC64;

Query March 13.7%; Score 639.5; DB 2; Length 767;
Best Local Similarity 26.6%; Pred. No. 1.5e-35;
Matches 170; Conservative 71; Mismatches 215; Indels 183; Gaps 17;

Qy 365 DCTWN-----IEVNNQHVYKRFKPYLLRPGVAGTCKPKYVEINGERKYC--GERSQ 415
Db 150 DCPYGDDEKNCATPTPTPCQWYCSYY-----TCYGVQICNGVQDDCPYGDDE 200
Qy 416 FVYTSNKNKTVRPHSDQSTDTGFLAEYLSDSPDPCQFTCRG----- 462
Db 201 NCATKTPSIFTCQWYC--SYTSTCIYGVQI-----CNGVQDAYGDDEKNCATKTPSI 251
Qy 463 -----RCIKRELKRDGMADCTDHSDELNCSQDAHQFTCK-----NKRCKP 503
Db 252 PTCQLYCSYYTYTCLYAVQICNGVLDCPFVDERKCVIATSTPCQIYCKDPMFDYTCIY 311
Qy 504 LFWVCDVNDGDSNDEQGC-----SCPAQTFRCNSN-----GKCLSKSQCNCKD 548
Db 312 AYQWCDGVRQCYVDDERNCVATATTTATTTSPPTCOIYCMNPMYYTYTCLYAVQMGQV 371
Qy 549 DCGDGSDEASC-PKVNVTCTKATYRCNLGLCLSKNPECDGKEDGSDSDKDC----- 602
Db 372 QCYYGDDELNCDTFTTAYCEK--RCGSSVSCVLSQWCDGVDPCPYGBDEMSCVSLYP 428
Qy 603 ----- 602
Db 429 ADPOLQVYTSVSAMLPVCSDDYMNDFGRPAQDFGNGSSYNRYDTLMSPIYAPNGYFKL 488
Qy 603 -----DCGLRSFTROARVVGTDADGEEMFQVSL 632

```

```

Db      489 YSGYMRSKFYTSVQYSSCYSGNVVSLHCISCGVSNNSLVSRIVGTFANLGNMPCQVNL 548
Qy      633 HALGQGHICGASLTISPMLVSAAHCYIDRGFRYSDDPTQWTAFLGLHDOSORSAFQER 692
Db      549 QYI-TGYLCGGSIIISPKMIVTAAHCYGS---YSSASGMRVPAG--TLTKPSYTNASAY 601
Qy      693 RLKRIISHPPFNDFTFDYDIALLELEKPAEYSSMVRPICLPDASHVPAGKAIWVTGMH 752
Db      602 FVERIIVHPGKSYTYNDIALMKLRDEITFGYTTQPVCLPNSGMFWBAGTTTWISGWS 661
Qy      753 TQYGTGALLIÖKEIRIVINQTCEN--LBPQITPRMVCVFLSGGVDSQGDSSGGLS 810
Db      662 TYEGGSVSTYLYQYALPLIDSNVCNOSYVYNGQITSSMICAGYLSGVDTCCQGDSSGPLV 721
Qy      811 SVEADGRIFGAGVYVSWGDCAGRNKPGVYTRLPPLFRDMI 849
Db      722 N-KRNGTWMVLVGDTSKGDGCRANKPGVYGNVTTFLEMI 759

```

Search completed: November 29, 2004, 08:32:31
 Job time : 150.315 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 29, 2004, 08:19:13 ; Search time 113.685 Seconds

(without alignments)
3456.743 Million cell updates/sec

Title: US-09-936-333-5

Perfect score: 3789
Sequence: 1 MAERVYMLPPRARSLSKFV.....PGVYTRLPFRMDIKENTGV 683

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : 1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3760	99.2	855	1 ST14_HUMAN	O95Y66 homo sapien
2	3209	84.7	855	2 BAC35125	P56677 mus muscu
3	3209	84.7	855	2 BAC35125	Bac35125 mus muscu
4	3197	84.4	855	2 Q9JVC1	Q9J117 ratu
5	2379	62.8	422	2 Q8WVCI	Q8WVCI homo sapien
6	2280.5	60.2	845	2 Q6GR54	Q6GR54 xenopus lae
7	2254.5	59.5	845	2 Q6DGR1	Q6DGR1 xenopus lae
8	1543	40.7	663	2 Q6DEV0	Q6DEV0 xenopus tro
9	1017.5	26.9	572	2 Q7RTY8	Q7RTY8 homo sapien
10	1017.5	26.9	572	2 BAD18401	BAD18401 homo sapi
11	1015.5	26.8	572	2 Q8BIR6	Q8BIR6 mus muscu
12	1014	26.8	799	2 Q6PF94	Q6PF94 mus muscu
13	1014	26.8	799	2 AAH57674	AAH57674 mus muscu
14	1014	26.8	811	1 TMS6_MOUSE	Q8B10 mus muscu
15	1007	26.6	811	1 TMS6_MOUSE	Q8B10 mus muscu
16	1005	26.5	802	2 AA088764	AA088764 homo sapi
17	1005	26.5	802	2 AA088764	AA088764 homo sapi
18	985	26.0	824	2 Q6ICC2	Q6ICC2 homo sapien
19	985	26.0	824	2 CAG30332	CAG30332 homo sapi
20	718.5	19.0	1034	1 ENTX_PIG	P88074 sus scrofa
21	717.5	18.9	855	2 Q7Z410	Q7Z410 homo sapien
22	717.5	18.9	1059	2 Q7Z411	Q7Z411 homo sapien
23	703.5	18.6	1035	1 ENTX_BOVIN	P98072 bos taurus
24	694	18.3	1042	1 COR1_HUMAN	Q9Y545 homo sapien
25	690.5	18.2	1111	2 Q80VY4	Q80VY4 rattus norv
26	688	18.0	1019	1 ENTX_HUMAN	P88073 homo sapien
27	683.5	17.9	1069	1 ENTX_MOUSE	Q94319 mus muscu
28	677.5	17.5	1113	1 COR1_MOUSE	Q94319 mus muscu
29	663.5	17.0	767	2 Q6DGR2	Q6DGR2 xenopus lae
30	644.5	17.0	722	2 Q6NUP5	Q6NUP5 xenopus lae
31	644	17.0	722	2 Q6NUP5	Q6NUP5 xenopus lae

32	644	17.0	722	2 AAH68636	AAH68636 xenopus 1
33	632.5	16.7	680	2 Q868H7	Q868H7 branchiole
34	622	16.4	581	2 Q9XZM7	Q9XZM7 strongyloce
35	619.5	16.3	680	2 Q868H5	Q868H5 branchiole
36	608	16.0	688	2 Q868H6	Q868H6 branchiole
37	601	15.9	490	2 Q7TND4	Q7TND4 mus muscu
38	600	15.8	490	2 TMS2_MOUSE	Q9J148 mus muscu
39	598.5	15.8	490	2 Q6P7D7	Q6P7D7 rattus norv
40	598.5	15.8	490	2 AAH61712	AAH61712 rattus no
41	598	15.8	490	2 Q920K3	Q920K3 rattus norv
42	593	15.7	688	2 Q868H4	Q868H4 branchiole
43	588	15.5	422	1 DES1_HUMAN	Q9U152 homo sapien
44	588	15.5	423	2 Q6UW31	Q6UW31 homo sapien
45	588	15.5	423	2 AA089376	AA089376 homo sapi
46	584.5	15.4	701	2 Q9J3S9	Q9J3S9 rattus norv
47	584.5	15.4	868	2 Q9Y1V3	Q9Y1V3 polyandroca
48	581	15.3	703	2 Q7RTY7	Q7RTY7 homo sapien
49	578	15.3	704	2 Q8CHN8	Q8CHN8 rattus norv
50	577.5	15.2	704	1 CRAR_MOUSE	P98064 mus muscu
51	576.5	15.2	1524	2 Q91674	Q91674 xenopus lae
52	571	15.1	676	2 Q6IR15	Q6IR15 rattus norv
53	571	15.1	676	2 Q6DUU6	Q6DUU6 cyprinus ca
54	565	14.9	699	1 CRAR_HUMAN	P48740 h complem
55	558.5	14.7	492	1 TMS2_HUMAN	O15393 homo sapien
56	558.5	14.7	492	2 Q96T73	Q96T73 homo sapien
57	558.5	14.7	492	2 AAH51839	AAH51839 homo sapi
58	557.5	14.7	423	2 Q8BM10	Q8BM10 mus muscu
59	557	14.7	681	2 Q7ZT70	Q7ZT70 lampetra ja
60	553	14.6	698	2 Q6GEP9	Q6GEP9 xenopus lae
61	550	14.5	307	2 Q6ZND6	Q6ZND6 homo sapien
62	550	14.5	307	2 BAD18439	BAD18439 homo sapi
63	549.5	14.5	638	2 Q8R0P5	Q8R0P5 mus muscu
64	547	14.4	453	2 Q812A6	Q812A6 mus muscu
65	546	14.4	453	1 TMS3_MOUSE	Q8K10 mus muscu
66	545.5	14.4	638	1 KAL_MOUSE	P26262 mus muscu
67	543.5	14.3	453	2 Q6ZMC3	Q6ZMC3 homo sapien
68	543.5	14.3	453	2 AAQ88823	AAQ88823 homo sapi
69	543.5	14.3	453	2 BAD18806	BAD18806 homo sapi
70	541.5	14.3	417	2 Q8VH14	Q8VH14 rattus norv
71	539	14.2	707	2 Q8QGV0	Q8QGV0 cyprinus ca
72	535.5	14.1	279	2 Q9QZ74	Q9QZ74 rattus norv
73	533	14.1	454	1 TMS3_HUMAN	P57727 homo sapien
74	528.5	13.9	617	2 Q8UJ51	Q8UJ51 triakis scy
75	526.5	13.9	618	1 KAL_RAT	P14222 rattus norv
76	525.5	13.9	417	2 Q8VH11	Q8VH11 mus muscu
77	525.5	13.9	417	2 Q8VH18	Q8VH18 mus muscu
78	523.5	13.8	279	2 Q7TNX3	Q7TNX3 mus muscu
79	521	13.8	600	2 Q7ZTR2	Q7ZTR2 xenopus lae
80	518	13.7	603	1 CPA1_MOUSE	Q61139 mus muscu
81	518	13.7	604	2 CPA1_RAT	Q9WUJ3 rattus norv
82	515.5	13.6	698	2 Q9PU71	Q9PU71 xenopus lae
83	513.5	13.6	537	2 Q9BYE1	Q9BYE1 homo sapien
84	513.5	13.6	581	2 Q9BYE2	Q9BYE2 homo sapien
85	513	13.5	455	1 TMS5_MOUSE	Q9E04 mus muscu
86	513	13.5	471	2 Q8CFE0	Q8CFE0 mus muscu
87	512	13.5	666	2 Q6O1O9	Q6O1O9 gallus gall
88	512	13.5	686	2 AA873179	AA873179 gallus ga
89	511.5	13.5	648	1 HATT_HUMAN	Q6U235 homo sapien
90	510	13.5	645	2 Q7PWE4	Q7PWE4 anopheles g
91	508.5	13.4	445	2 Q8CJ17	Q8CJ17 rattus norv
92	508.5	13.4	467	2 Q967X8	Q967X8 panulirus a
93	507.5	13.4	250	2 Q9V514	Q9V514 drosophila
94	507.5	13.4	438	2 Q6ZWK6	Q6ZWK6 homo sapien
95	507.5	13.4	438	2 BAC85495	BAC85495 homo sapi
96	506.5	13.4	455	2 Q8CDR0	Q8CDR0 mus muscu
97	506	13.4	613	2 Q03711	Q03711 xenopus lae
98	506	13.4	719	2 Q6DJ90	Q6DJ90 xenopus tro
99	504.5	13.3	558	2 Q86VY4	Q86VY4 homo sapien
100	504	13.3	790	1 PLAN_PIG	P06867 sus scrofa

ALIGNMENTS

RESULT 1
ST14_HUMAN STANDARD; PRT; 855 AA.
ID TISSUE=Milk; PubMed=9303582; PubMed=10373425;
AC Q9V5T6; Q9B501; Q9H3S0; Q9H3B6; Q9HCA3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE Suppressor of tumorigenicity 14 (EC 3.4.21.-) (Matrilysin (Membrane-type serine protease 1) (MT-SP1) (Prostasin) (Serine protease TADG-15) (Tumor associated differentially-expressed gene-15 protein).
GN Name=ST14; Synonyms=PRSS14, SNCL19, TADG15;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9303581; PubMed=10373424;
RA Lin C.Y., Anders J., Johnson M., Sang Q.A., Dickson R.B.;
RT "Molecular cloning of cDNA for matrilysin, a matrix-degrading serine protease with trypsin-like activity";
RL J. Biol. Chem. 274:18231-18236 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99432178; PubMed=10500122;
RA Takeuchi T., Shuman M.A., Craik C.S.;
RT "Reverse biochemistry: use of macromolecular protease inhibitors to dissect complex biological processes and identify a membrane-type serine protease in epithelial cancer and normal tissue";
RL Proc. Natl. Acad. Sci. U.S.A. 96:11054-11061 (1999).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Prostate;
RA Yamaguchi N., Mitsui S.;
RT "Molecular cloning of a novel transmembrane serine protease expressed in human prostate";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Tanimoto H., Underwood L.J., Wang Y., Shigemasa K., Parmley T.H., O'Brien T.J.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC TISSUE=Blood; and Muscle;
RX MEDLINE=22386257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F., Diatchenko L., Marusik K., Farmer A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein W.J., Umed T.B., Tomihataki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullenbach S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fehey J., Helton E., Kesteman M., Madan A., Rodriguez S., Sanchez A., Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmitt J., Myers R.M., Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E., Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [6]
RP SEQUENCE OF 340-664 FROM N.A.
RA Cao J., Fan W., Zheng S.;
RT "Genomic analysis of a novel human serine protease SNCL19";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [7]

RP CHARACTERIZATION.
RC TISSUE=Milk;
RX MEDLINE=9303582; PubMed=10373425;
RA Lin C.Y., Anders J., Johnson M., Dickson R.B.;
RT "Purification and characterization of a complex containing matrilysin and a Kunitz-type serine protease inhibitor from human milk";
RL J. Biol. Chem. 274:18237-18242 (1999).
CC -1- FUNCTION: Degrades extracellular matrix. Proposed to play a role in breast cancer invasion and metastasis. Exhibits trypsin-like activity as defined by cleavage of synthetic substrates with Arg or Lys as the P1 site.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
CC -1- SIMILARITY: Belongs to peptidase family S1.
CC -1- SIMILARITY: Contains 2 CUB domains.
CC -1- SIMILARITY: Contains 4 LDL-receptor class A domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation- the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF118224; AAD42765.2; -;
DR EMBL; AF133086; AAF00109.1; -;
DR EMBL; AB030036; BAB20376.1; -;
DR EMBL; AF057145; AAG15395.1; -;
DR EMBL; BC005826; AAH05826.2; -;
DR EMBL; BC030532; AAH30532.1; -;
DR EMBL; AF283256; AAG13949.1; -;
DR PDB; 1EAW; X-ray; A/C=615-855.
DR MEROPS; S01.302; -;
DR Genew; HGNC:11344; ST14.
DR MIM; 606797; -;
DR GO; GO:0005887; C:integral to plasma membrane; TAS.
DR GO; GO:0008236; F:serine-type peptidase activity; TAS.
DR GO; GO:0006508; P:proteolysis and peptidolysis; TAS.
DR InterPro; IPR000859; CUB.
DR InterPro; IPR002125; LDL_receptor_A.
DR InterPro; IPR001314; Peptidase_S1A.
DR InterPro; IPR009003; Pept_Ser_Cys.
DR Pfam; PF00431; CUB; 2.
DR Pfam; PF00057; Ldl_recept_a; 4.
DR Pfam; PF00089; Trypsin; 1.
DR PRINTS; PR00222; CHYMOTRYPSIN.
DR PRINTS; PR00261; LDLRECEPTOR.
DR SMART; SM00192; LDLa; 4.
DR SMART; SM00020; TRYP_SPC; 1.
DR PROSITE; PS01180; CUB; 2.
DR PROSITE; PS01209; LDLRA_1; 2.
DR PROSITE; PS0068; LDLRA_2; 4.
DR PROSITE; PS00240; TRYP_SIN_DOM; 1.
DR PROSITE; PS00134; TRYP_SIN_HIS; 1.
DR PROSITE; PS00135; TRYP_SIN_SER; 1.
KW 3D-structure; Glycoprotein; Hydrolase; Repeat; Serine protease;
KW Signal-anchor; Transmembrane.
FT DOMAIN 1 55 Cytoplasmic (Potential).
FT TRANSMEM 56 76 Signal-anchor for type II membrane protein (Potential).
FT FT Extracellular (Potential).
FT DOMAIN 77 855 CUB 1.
FT DOMAIN 214 334 CUB 2.
FT DOMAIN 452 487 LDL-receptor class A 1.
FT DOMAIN 487 524 LDL-receptor class A 2.
FT DOMAIN 524 560 LDL-receptor class A 3.
FT DOMAIN 566 603 LDL-receptor class A 4.
FT DOMAIN 615 854 Serine protease.
FT ACT_SITE 656 Charge relay system (By similarity).
FT ACT_SITE 711 Charge relay system (By similarity).
FT ACT_SITE 805 Charge relay system (By similarity).

```

FT CARBOHYD 109 109 N-linked (GlcNAc... ) (Potential)
FT CARBOHYD 302 302 N-linked (GlcNAc... ) (Potential)
FT CARBOHYD 485 485 N-linked (GlcNAc... ) (Potential)
FT CARBOHYD 772 772 N-linked (GlcNAc... ) (Potential)
FT CONFLICT 381 381 A -> S (in Ref. 4)
FT CONFLICT 674 674 R -> V (in Ref. 3)
SQ SEQUENCE 855 AA; 94769 MW; 26143132C01F99C9 CRC64;

Query Match 99.2%; Score 3760; DB 1; Length 855;
Best Local Similarity 99.4%; Pred. No. 3,8e-236;
Matches 679; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAERVVMLPPRASLSKSFVTVSVVAFPTDSTQVORTDSCSGFLHARGVLEKRFPTPG 60
DB 173 MAERVVMLPPRASLSKSFVTVSVVAFPTDSTQVORTDSCSGFLHARGVLEKRFPTPG 232
QY 61 FPDSPYPAHARQNALRGDADSVLSLTFRSPDLASCDERGSGLVTVYNTLSPMEPHALVQ 120
DB 233 FPDSPYPAHARQNALRGDADSVLSLTFRSPDLASCDERGSGLVTVYNTLSPMEPHALVQ 292
QY 121 LCGTTPSYNLTFRSSQVNLITLTNTERRHPGEAFPLPMSSCGGLRKAQGTEN 180
DB 293 LCGTTPSYNLTFRSSQVNLITLTNTERRHPGEAFPLPMSSCGGLRKAQGTEN 352
QY 181 SPYRGHYPPNIDCTWNIENVNNOHVKKRFKFLLEPRACGTCPCDQYVEINKEKCGE 240
DB 353 SPYRGHYPPNIDCTWNIENVNNOHVKKRFKFLLEPRACGTCPCDQYVEINKEKCGE 412
QY 241 RSQFVTVSNNSKNTVRFHSDSYTDTGFLAEVLSYSSDSPCPQGTCTGRCIRKELNCD 300
DB 413 RSQFVTVSNNSKNTVRFHSDSYTDTGFLAEVLSYSSDSPCPQGTCTGRCIRKELNCD 472
QY 301 GMADCTHSDSLNCSGAGHFTCKNCKRCLFWVCDSDVNDGNSDPSGSCCAQGTFR 360
DB 473 GMADCTHSDSLNCSGAGHFTCKNCKRCLFWVCDSDVNDGNSDPSGSCCAQGTFR 532
QY 361 SNGKCLSKSQCGKNGDGDSDSDEASCPKVVVCTKTKTYCLNGTCLSKGNPECDGED 420
DB 533 SNGKCLSKSQCGKNGDGDSDSDEASCPKVVVCTKTKTYCLNGTCLSKGNPECDGED 592
QY 421 CSDSDSEKDCDCCGLRSFTQARVVGTDADGEPMQVSLHALQGHICGASLISPMVLV 480
DB 593 CSDSDSEKDCDCCGLRSFTQARVVGTDADGEPMQVSLHALQGHICGASLISPMVLV 652
QY 481 SAACHYIDDRGFRSDPTQMTAFGLHDSQSRAPGVQERLKLIIHPFNDDTFDYDI 540
DB 653 SAACHYIDDRGFRSDPTQMTAFGLHDSQSRAPGVQERLKLIIHPFNDDTFDYDI 712
QY 541 ALLELEKPAEYSSNVRFICLPDASHVFPAGKAIWVGHTQYGGTGALLIQKEIRVYN 600
DB 713 ALLELEKPAEYSSNVRFICLPDASHVFPAGKAIWVGHTQYGGTGALLIQKEIRVYN 772
QY 601 QTTCENTLIPQOITRRMPCVFLSGGVDSQCGDSGGPLSSVEADGRIFQAGVSVMGDCAQ 660
DB 773 QTTCENTLIPQOITRRMPCVFLSGGVDSQCGDSGGPLSSVEADGRIFQAGVSVMGDCAQ 832
QY 661 RNKPGVYTRLPFRDMIKENTGV 683
DB 833 RNKPGVYTRLPFRDMIKENTGV 855

RESULT 2
ST14_MOUSE STANDARD; PRT; 855 AA.
AC P56677;
DT 15-JUL-1999 (Rel. 38, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 05-JUN-2004 (Rel. 44, Last annotation update)
DE Suppressor of tumorigenicity 14 (EC 3.4.21.-) (Epithin).
GN Name=St14; Synonym=Prsbl14;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

```

OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C.B.17SCID; TISSUE=Thymus;
RX MEDLINE=99216440; PubMed=10199918;
RA Kim M.G., Chen C., Lyu M.S., Cho E.G., Park D., Kozak C.,
RA Schwartz R.H.;
RT "Cloning and chromosomal mapping of a gene isolated from thymic
RT stromal cells encoding a new mouse type II membrane serine protease,
RT epithin, containing four LDL receptor modules and two CUB domains.";
RL Immunogenetics 49:420-428(1999).
RN [2]
RP REVISIONS TO 23; 321; 325; 343; 409-410 AND C-TERMINUS.
RC STRAIN=C.B.17SCID; TISSUE=Thymus;
RA Kim M.G., Chen C., Cho E.G., Park D., Schwartz R.H.;
RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast tumor;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold B.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mulhally S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Vallatton D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Faley J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley K.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W.,
RA Butcherfield Y.S.N., Krzywinski M.I., Skalka A., Smalins D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences".
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
CC -1- TISSUE SPECIFICITY: Highly expressed in intestine, kidney, lung,
CC and thymus. Not expressed in skeletal muscle, liver, heart, testis
CC and brain.
CC -1- SIMILARITY: Belongs to peptidase family S1.
CC -1- SIMILARITY: Contains 2 CUB domains.
CC -1- SIMILARITY: Contains 4 LDL-receptor class A domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi-sib.ch).
CC -----
DR EMBL; AF042822; AA002220.3; -
DR EMBL; BC005496; AA005496.1; -
DR HSSP; P00760; IEZX.
DR MEROPS; S01.302; -.
DR MGD; MGI:1338881; St14.
DR GO; GO:0005576; C:extracellular; IDA.
DR GO; GO:0019897; C:extrinsic to plasma membrane; IDA.
DR GO; GO:0008236; F:serine-type peptidase activity; IDA.
DR InterPro; IPR000859; CUB.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR InterPro; IPR009003; Pept_Ser_Cys.
DR Pfam; PF00431; CUB; 2.
DR Pfam; PF00057; LDL_recept-_a; 4.
DR Pfam; PF00089; Trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.

```

DR	PRINTS, P000261, LDLRECEPTOR.	
DR	SMART, SM00042; CUB; 2.	
DR	SMART, SM00192; LDLA; 4.	
DR	SMART, SM00020; TRYP_SPC; 1.	
DR	PROSITE; PS00180; CUB; 2.	
DR	PROSITE; PS01009; LDLA_1; 2.	
DR	PROSITE; PS05068; LDLA_2; 4.	
DR	PROSITE; PSS0240; TRYPSIN_DOM; 1.	
DR	PROSITE; PSS00134; TRYPSIN_HIS; 1.	
DR	PROSITE; PS00135; TRYPSIN_SER; 1.	
KW	Glycoprotein, Hydrolase; Repeat; Serine protease; Signal-anchor; Transmembrane.	
FT	DOMAIN	1
FT	TRANSMEM	56
FT		76
FT	DOMAIN	77
FT	DOMAIN	214
FT	DOMAIN	340
FT	DOMAIN	451
FT	DOMAIN	489
FT	DOMAIN	523
FT	DOMAIN	565
FT	DOMAIN	615
FT	ACT_SITE	656
FT	ACT_SITE	711
FT	ACT_SITE	711
FT	ACT_SITE	805
FT	CARBOHYD	107
FT	CARBOHYD	302
FT	CARBOHYD	365
FT	CARBOHYD	421
FT	CARBOHYD	489
FT	CARBOHYD	772
FT	SEQUENCE	855 AA; 94654 MM; 4F10E84DA2A146D5 CRC64;

Query Match	84.7%;	Score 3209;	DB 1;	Length 855;
Best Local Similarity	83.2%;	Pred. No. 2.3e-217;		
Matches 568;	Conservative 59;	Mismatches 56;	Indels 0;	Gaps 0;

Qy	1	MAEERVYMLPPRRASLSKSVYVTSVAFAPTDSDKTVQRODNCSPSGHLARAGVLMRFPTTG	60
Db	173	MAEERVYVLLPPRRRAALKSKFVLTSSVAFAFIDPDMALQRODNCSPALHAGAVIRFPTTG	2322
Qy	61	FPDSPYPAHARCOMALRGDADSVLSITERSPDLASCDRGSDLYTVYNTLSPMEPAHAYQ	120
Db	233	FPNSPYPAAHARCOMVLRGADSVLSITERSFDVACDGHGSDLYTVYDLSLSPMEPAHAYR	2922
Qy	121	LCGTTPSPYVLTTHSSQNVLLTTLTNTBRHARGFAAFPOLPMSSCGRLRAQSTEN	180
Db	293	LCGTFSBPNVLTPLSSQNVLTTLTNTDRRRPGEAFPOLPMSSCGGLSDTQSTFS	3522
Qy	181	SPYPPGHYPENIDCTNNIEVPPNNOHVKAFKFEYLLBERRACGTCPXDYVEINERKCGE	240
Db	353	SPYPPGHYPENINCTNNIKVPPNNRVKVRFKFLYLVDPNVVSGCTDYEINERKCGE	412
Qy	241	RSQFPVYTSNSNKTIVAFPHSDQSYTDTGFLAETLSLSDSDPCCGQFTCRGCTRKELRCD	300
Db	413	RSQFPVVSNSSKITVAFPHSDSYTDTGFLAETLSLSDSDPCGMFMCKTGCTRIKELRCD	4722
Qy	301	GMADCTDHSDELNCSCDAGHQFTCKNKKFCKPLFWVCDSDVNCGNSDEOGSCSPAQTFRC	360
Db	473	GMADCDYDGERYCRNATHQFTCKNQPCKPLFWVCDSDVNCGNSDEBGSCSPAQSFRC	5322
Qy	361	SNKCKLSKSGQCNCKADCCBGSDEASCPKRVNVYTCFKTFFYRLCLNGLCLSKGNPECDKED	420
Db	533	SNKCKLPQSQCKNGKDNCCDGSDEASCDSDVNVVSCFKTYRCCQNGCLCSKGNPECCGKTD	5922
Qy	421	CSGSGSEKCCDCGLRSFTQCARVVGCTADDEGEMPMOYSLHALCGHICGASLISPMWLY	480
Db	593	CSGSGSEKCCDCGLRSFTQCARVVGCTADDEGEMPMOYSLHALCGHICGASLISPMWLY	6522
Qy	481	SAACHYIIDRGFRYSDPTQWTAFGLGHQDSORSAFGVGERLRKLIISHPFNDETFDYDI	540
Db	653	SAACHCQDDKNFTYSYTYMTWTFGLGLQDSKSSAGVDELKTKLIIITPSSNFTTFDYDI	7122

OY		541	ALTELEKAEASWNRPlCtLPASAVPAPGKAiWTGTGHTQGGTGAIILOKGEIRVIn	600
Dd		713	ALLELEKEVESISTVRPiCLPATVIFPPGKAiWGTGHTKEGGGAILLOKGEIRVIN	772
OY		601	OTTCENLPLPOOITTPRPMCMVGFLSGGVDSOCQDSSGGPGLSSVEADGRIFQAQVWSMGDCAQ	660
Dd		773	OTTCEDLMPQQITTRPMCMVGFLSGGVDSOCQDSSGGPGLSSAEXDGNRFQAQVWSMGSCAQ	832
OY		661	RNKPGVYTRLPLFRDWIKENTGV	693
Dd		833	RNKPGVYTRLPLFVRDWIKENHTGV	855
RESULT 3				
BAC35125				
ID	BAC35125;	PRELIMINARY;	PRT;	855 AA.
AC	BAC35125;			
DT	14-APR-2004 (TREMBLrel. 27, Created)			
DT	14-APR-2004 (TREMBLrel. 27, Last sequence update)			
DT	14-APR-2004 (TREMBLrel. 27, Last annotation update)			
DE	clone:pe30004B06 product:suppression of tumorigenicity 14 (colon carcinoma), full insert sequence.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCB1_TaxId=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=Kidney;			
RX	MEDLINE=22354683; PubMed=12466651;			
RA	The FANTOM Consortium,			
RA	the RIKEN Genome Exploration Research Group Phase I & II Team,			
RT	"Analysis of the mouse transcriptome based on functional annotation of			
RL	60,770 full-length cDNAs."			
RN	Nature 420:563-573(2002).			
[2]				
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=Kidney;			
RX	MEDLINE=21085660; PubMed=11217851;			
RA	RIKEN FANTOM Consortium;			
RA	"Functional annotation of a full-length mouse cDNA collection.";			
RL	Nature 409:685-690(2001).			
[3]				
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=Kidney;			
RX	MEDLINE=99279253; PubMed=10349636;			
RA	Carninci P., Hayashizaki Y. ;			
RT	"High-efficiency full-length cDNA cloning.";			
RL	Meth. Enzymol. 303:19-44(1999).			
[4]				
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=Kidney;			
RX	MEDLINE=20499374; PubMed=11042159;			
RA	Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,			
RA	Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y. ;			
RT	"Normalization and subtraction of cap-trapper-selected cDNAs to			
RT	prepare full-length cDNA libraries for rapid discovery of new genes.";			
RL	Genome Res. 10:1617-1630(2000).			
[5]				
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=Kidney;			
RX	MEDLINE=20530913; PubMed=11076861.			
RA	Shibata K., Itoh M., Atzawa K., Nagaoaka S., Saeki N., Carninci P.,			
RA	Kono H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,			
RA	Sun H., Ishii Y., Nakamura S., Hazama M., Nishino T., Harada A.,			
RA	Yanamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashtwagi K.,			
RA	Fujikawa S., Inoue K., Togawa Y., Izawa M., Ohara E., Wachihi M.,			
RA	Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,			
RA	Okazaki Y., Muramatsu M., Inoue Y., Kita A., Hayashizaki Y. ;			
RT	"RIKEN integrated sequence analysis (RISA) system-384-format			
RT	sequencing pipeline with 384 multicapillary sequencer.";			

RL Genome Res. 10:1757-1771 (2000).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Kidney;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagata T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA Hori F., Imocani K., Iehi Y., Itoh M., Kagawa I., Kanakawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito D., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Saeki R., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Yamamatsu M., Hayashizaki Y.,
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK052738; BAC35125.1; -
SQ SEQUENCE 855 AA; 94654 MW; 4F10E84DA2146DD5 CRC64;

Query Match 84.4%; Score 3209; DB 2; Length 855;
Best Local Similarity 83.2%; Pred. No. 2.3e-217;
Matches 568; Conservative 59; Mismatches 56; Indels 0; Gaps 0;

QY 1 MAERVVMLEPPARSLKSFVVTSVVAFPDSTQVORTDNCSCFGLHARGVELMRFTTPG 60
DB 173 MAVERVVTLPARRALKSFVLTSVVAFPIDPRMLQRTDNCSCFGLHARGAAYTRFTTPG 232
QY 61 PPDSPYPAHARCOMALRGDADSVLSLTFRSFDLASCDERSGLVTVYNTLSPEMHPALVQ 120
DB 233 PPNSPYPAHARCOMVLRGDADSVLSLTFRSFDVAPCDHSGSDLVTVYDLSLSPMEHPAVVR 292
QY 121 LCGTTPPSYNTLTFHSSQNVLLITLITNTERRHGFEATFQLPRMSSCGGLRLKAQGTEN 180
DB 293 LCGTSPSYNTLTFSSQNVFLVTLITNDRHPGEATFQLPRMSSCGGLSLPTQGTFS 352
QY 181 SPYYPGHYPNIDCTWNIIEVNNQHVKRFKFFYLPRRAAGTCTPKDYVINGEKYCGE 240
DB 353 SPYYPGHYPNIDCTWNIIEVNNQHVKRFKFFYLPRRAAGTCTPKDYVINGEKYCGE 412
QY 241 RSQFVTVSNKNTVRPHSDSYDTDTGFLAELYSYSSDPCPGQFTCTGRCIRKELRCD 300
DB 413 RSQFVTVSNKNTVRPHSDSYDTDTGFLAELYSYSSDPCPGQFTCTGRCIRKELRCD 472
QY 301 GMADCTHSDLELNCSCAGHOFCTCKNRCRKLFWVCSVNDCCGNSDEGSCSCPAQTRFC 360
DB 473 GMADCPYSDRYRCNATHQFTCKNQCKPLFWVCSVNDCCGNSDEGSCSCPAQTRFC 532
QY 361 SNGKCLSKSQCGKDCGDSDEASCPKVVVYCTKITYRCLNGLCLSKNPECDGKED 420
DB 533 SNGKCLSKSQCGKDCGDSDEASCPKVVVYCTKITYRCLNGLCLSKNPECDGKED 592
QY 421 CSDSDSEKDCDGLRSTFRQARVVGTDADGEMPMQVSLHALQGHICGASLISPMWLV 480
DB 593 CSDSDSEKDCDGLRSTFRQARVVGTDADGEMPMQVSLHALQGHICGASLISPMWLV 652
QY 481 SAACIYIDDRFRSDPTQWTAFLGHDGORSAPRGVORLKLIIHSFPNDFTFYDI 540
DB 653 SAACIYIDDRFRSDPTQWTAFLGHDGORSAPRGVORLKLIIHSFPNDFTFYDI 712
QY 541 ALLELEKPAEYSSWVRPCLPDASHVPFAGRAIWTGNGHTQYGGTGLIIQKEIRIYN 600
DB 713 ALLELEKPAEYSSWVRPCLPDASHVPFAGRAIWTGNGHTQYGGTGLIIQKEIRIYN 772
QY 601 QTTCENTLLPQOITPRMCMVGLSGGVDSQCGDSGGPLSSVADGRIFQAGVSWGDCGQ 660
DB 773 QTTCENTLLPQOITPRMCMVGLSGGVDSQCGDSGGPLSSVADGRIFQAGVSWGDCGQ 832
QY 661 RNKPGVYTRLPFRDWMKENTGV 683
DB 833 RNKPGVYTRLPFRDWMKENTGV 855

RESULT 4

09JUL7
ID 09JUL7 PRELIMINARY; PRT; 855 AA.
AC 09JUL7
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 05-JUN-2004 (TrEMBLrel. 27, Last annotation update)
DE Membrane bound serine protease (Membrane bound arginine specific
DE serine protease).
GN Name=MBSP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar; TISSUE=Jejunum;
RA MEDLINE=21458307; Pubmed=11573963;
RA Sacomi S., Yamaoka Y., Tsuzuki S., Hitomi Y., Iwanaga T., Fushiki T.,
RT "A role for membrane-type serine protease (MT-SPI) in intestinal
RT epithelial turnover".
RL Biochem. Biophys. Res. Commun. 287:995-1002 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar; TISSUE=Duodenum;
RA Inoue H., Takahashi K., Kishi K.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Belongs to peptidase family 11.
DR EMBL; AB037898; BAB03502.1; -
DR EMBL; AB049189; BAB13765.1; -
DR PIR; JC7731; JC7731.
DR HSSP; P00760; 1EZX.
DR MEROPS; S01.302; -
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000859; CUB.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001254; peptidase_S1.
DR InterPro; IPR001314; peptidase_S1A.
DR InterPro; IPR009003; Pept_Ser_Cys.
DR Pfam; PF00431; CUB; 2.
DR Pfam; PF00057; Ldl_recept_a; 4.
DR Pfam; PF00089; Trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00261; LDLRECEPTOR.
DR PROSITE; PS01180; CUB; 2.
DR PROSITE; PS01209; LDLRA_1; 2.
DR PROSITE; PS00068; LDLRA_2; 4.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Protease; Serine protease.
SQ SEQUENCE 855 AA; 94955 MW; 35806B7ECFCF03D CRC64;

Query Match 84.4%; Score 3197; DB 2; Length 855;
Best Local Similarity 83.0%; Pred. No. 1.6e-216;
Matches 567; Conservative 59; Mismatches 57; Indels 0; Gaps 0;

QY 1 MAERVVMLEPPARSLKSFVVTSVVAFPDSTQVORTDNCSCFGLHARGVELMRFTTPG 60
DB 173 MAVERVVTLPARRALKSFVLTSVVAFPIDPRMLQRTDNCSCFGLHARGAAYTRFTTPG 232
QY 61 PPDSPYPAHARCOMALRGDADSVLSLTFRSFDLASCDERSGLVTVYNTLSPEMHPALVQ 120
DB 233 PPNSPYPAHARCOMVLRGDADSVLSLTFRSFDVAPCDHSGSDLVTVYDLSLSPMEHPAVVR 292
QY 121 LCGTTPPSYNTLTFHSSQNVLLITLITNTERRHGFEATFQLPRMSSCGGLRLKAQGTEN 180
DB 293 LCGTSPSYNTLTFSSQNVFLVTLITNDRHPGEATFQLPRMSSCGGLSLPTQGTFS 352
QY 181 SPYYPGHYPNIDCTWNIIEVNNQHVKRFKFFYLPRRAAGTCTPKDYVINGEKYCGE 240
DB 353 SPYYPGHYPNIDCTWNIIEVNNQHVKRFKFFYLPRRAAGTCTPKDYVINGEKYCGE 412

```
Dh 353 SPYYPGHPYNNINCTWNIKVPNNNNKVRKFLFYLVDPENIPVSGCTKDYVEINKEKCE 412
Qy 241 RSQFVTSNNSKLTIVRFRSDSYTDTGFLAEYLSYDSSDPGQFTCRGTGRKIKELRCD 300
Dh 413 RSQFVTSNNSKLTIVRFRSDSYTDTGFLAEYLSYDSSDPGQFTCRGTGRKIKELRCD 472
Qy 301 GMADCTHSDBELNCSGAGHQFTCKNFKCKPLFVVCDSVNDGDSNDQSCSPAQTFRC 360
Dh 473 GMADCPYSDERHRCNATHQFMCKNQFCKPLFVVCDSVNDGDSNDQSCSPAQTFRC 532
Qy 361 SNGKCLSKSQOCCNKDCCGSDSDASCPKVVVYCTHTYRCNLGLCLSKNPECDKED 420
Dh 533 SNGKCLSKSQOCCNKDCCGSDSDASCDNVNVAVCTYTRCQGLCLINKNPECDKED 592
Qy 421 CSDDSDKDCDCGSRFTKQARVVGCTDADGEMPMQVSLHALGQGHICGASLISPMVLV 480
Dh 593 CSDDSDKDCDCGSRFTKQARVVGCTDADGEMPMQVSLHALGQGHICGASLISPMVLV 652
Qy 481 SAACVYIDRGFRYSPTQWTAFLGLHDQSRAPGVQERLKRILSHPFNDFTPDYDI 540
Dh 653 SAACVYIDRGFRYSPTQWTAFLGLHDQSRAPGVQERLKRILSHPFNDFTPDYDI 712
Qy 541 ALLELEKPAEYSSWVRPCLPDASHVPRAGKAIWVTGHTQVGTGALLIQKEIRVIN 600
Dh 713 ALLELEKPAEYSSWVRPCLPDASHVPRAGKAIWVTGHTQVGTGALLIQKEIRVIN 772
Qy 601 QTTCEMLLPQOITPRMMCVGFLSGGVSCGDSGGLPSYEADGRIFQAGVSWGDCQAQ 660
Dh 773 QTTCEMLLPQOITPRMMCVGFLSGGVSCGDSGGLPSYEADGRIFQAGVSWGDCQAQ 832
Qy 661 RNKPGVYTRILPFDWIKENTGV 683
Dh 833 RNKPGVYTRILPFDWIKENTGV 855

RESULT 5
ID Q8WVC1 PRELIMINARY; PRT; 422 AA.
AC Q8WVC1;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE S14 protein (Fragment).
GN S14;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Colon;
RX MEDLINE=22388257; PubMed=12477932;
RA Krausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatzenko L., Marsina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Tohiyuki S., Carninci P., Prange C.J.,
RA Raha S.S., Loggiano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.U., Hulik S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalins D.E., Scherch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
```

```
RP SEQUENCE FROM N.A.
RC TISSUE=Colon;
RA Krausner R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Belongs to peptidase family S1.
DR EMBL; BC018146; AAI18146.1; -.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001314; Peptidase_S1.
DR InterPro; IPR009003; Pept_ser_Cys.
DR Pfam; PF00057; Ldl_recept_a; 4.
DR Pfam; PF00089; Trypsin_1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00261; LDLRECEPTOR.
DR SMART; SM00192; LDLA; 4.
DR SMART; SM00020; TRYP_Spc; 1.
DR PROSITE; PS01209; LDLA_1; 2.
DR PROSITE; PS50068; LDLA_2; 4.
DR PROSITE; PS50240; TRYPsin_DOM; 1.
DR PROSITE; PS00134; TRYPsin_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPsin_SER; 1.
KW Hydrolase; Protease; Serine protease.
FT NON TER
SQ SEQUENCE 422 AA; 46257 MW; 2C99875D1B58B319 CRC64;

Query Match 62.8%; Score 2379; DB 2; Length 422;
Best Local Similarity 100.0%; Pred. No.2.8e-159;
Matches 422; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 262 SYTDGFLAEYLSYDSSDPGQFTCRGTGRKIKELRCGMDCTHSDBELNCSGAGHQ 321
Dh 1 SYTDGFLAEYLSYDSSDPGQFTCRGTGRKIKELRCGMDCTHSDBELNCSGAGHQ 60
Qy 322 FTCKNFKCKPLFVVCDSVNDGDSNDQSCSPAQTFRCNSGKCLSKSQOCCNKDCCGDS 381
Dh 61 FTCKNFKCKPLFVVCDSVNDGDSNDQSCSPAQTFRCNSGKCLSKSQOCCNKDCCGDS 120
Qy 382 SDBASCPKVVVYCTHTYRCNLGLCLSKNPECDKEDCDSDSDKDCGSRFTKQAR 441
Dh 121 SDBASCPKVVVYCTHTYRCNLGLCLSKNPECDKEDCDSDSDKDCGSRFTKQAR 180
Qy 442 RVVGGTDADEGEMPMQVSLHALGQGHICGASLISPMVLVSAACVYIDRGFRYSPTQWT 501
Dh 181 RVVGGTDADEGEMPMQVSLHALGQGHICGASLISPMVLVSAACVYIDRGFRYSPTQWT 240
Qy 502 AFLGLHDQSRAPGVQERLKRILSHPFNDFTPDYIALLEKPAEYSSWVRPCLP 561
Dh 241 AFLGLHDQSRAPGVQERLKRILSHPFNDFTPDYIALLEKPAEYSSWVRPCLP 300
Qy 562 DASHVPRAGKAIWVTGHTQVGTGALLIQKEIRVINQTCENLLPQOITPRMMCVGF 621
Dh 301 DASHVPRAGKAIWVTGHTQVGTGALLIQKEIRVINQTCENLLPQOITPRMMCVGF 360
Qy 622 LSGGVSCGDSGGLPSYEADGRIFQAGVSWGDCQAQRNKGVTYTRILPFDWIKENT 681
Dh 361 LSGGVSCGDSGGLPSYEADGRIFQAGVSWGDCQAQRNKGVTYTRILPFDWIKENT 420
Qy 682 GV 683
Dh 421 GV 422

RESULT 6
ID Q6GR54 PRELIMINARY; PRT; 845 AA.
AC Q6GR54;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE S14-A-prov protein.
```


GN Name=act14-A-prov;
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
 OC Xenopodidae; Xenopus.
 CX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stepieton M., Soares M.B., Bonaldo M.F., Casavani T.L., Scheetz T.E.,
 RA Brownstein M.J., Ueda T.B., Toshitoki S., Carrinci P., Prange C.,
 RA Rata S.S., Loquellano N.A., Peters G.J., Adamson R.D., Mullaly S.J.,
 RA Bosak S.A., McGwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Feilting J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalka U., Smallos D.E., Scherch A., Schein J.E.,
 RA Jones S.J., Maitra M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RX MEDLINE=22341132; PubMed=12454917;
 RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
 RA Richardson P.,
 RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
 RT Initiative."
 RL Dev. Dyn. 225:384-391(2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RA Klein S., Strausberg R.,
 RL Submitted (May-2004) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: Belongs to peptidase family S1.
 DR EMBL: BC071077; AA071077.1; -.
 DR InterPro: IPR000859; CUB.
 DR InterPro: IPR002172; LDL_receptor_A.
 DR InterPro: IPR001254; Peptidase_S1.
 DR InterPro: IPR001314; Peptidase_S1A.
 DR InterPro: IPR009003; Pept_Ser_Cys.
 DR Pfam: PF00431; CUB; 2.
 DR Pfam: PF00057; Ldl_recept_a; 4.
 DR Pfam: PF00089; Trypsin; 1.
 DR PRINTS: PRO0722; CHYMOTRYPSIN.
 DR PRINTS: PRO0261; LDLRECEPTOR.
 DR SMART: SM00042; CUB; 2.
 DR SMART: SM00192; LDLA; 4.
 DR SMART: SM00020; Tryp_Spc; 1.
 DR PROSITE: PS01180; CUB; 2.
 DR PROSITE: PS01209; LDLRA_1; 2.
 DR PROSITE: PS00068; LDLRA_2; 4.
 DR PROSITE: PS02440; TRYPSIN_DOM; 1.
 DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.
 DR PROSITE: PS00135; TRYPSIN_SER; 1.
 KM HydroLase; Protease; Serine protease.
 SQ SEQUENCE 845 AA; 93593 MW; B917386C95BC73BD CRC64;

Db 164 ISELKLPVNRK---TFALDISLVAVPTDQRLARVKNSSCAVFLHSSNGVAKFFSPG 220
 Qy 61 PPDSPPYAHARCOMALRGDADSVLSLFRSFDLASCDERSDLVYNTLSPMEPHLVQ 120
 Db 221 PPDSPPYAHARCOMALRGDADSVLSLFRSFDLASCDERSDLVYNTLSPMEPHLVQ 280
 Qy 121 LCGTYPSPYALTFHSSONVLTITLTNTERRHGFEPQLPRMSCGRLKAQSTFN 180
 Db 281 LCGTYPSPYALTFHSSONVLTITLTNTERRHGFEPQLPRMSCGRLKAQSTFN 340
 Qy 181 SPYYPGHYPNIDCTWNIIEVPPNQHVPKPFYLLLEPRACGTCPPDYVEINGEKYCGE 240
 Db 341 SPYYPGHYPNIDCTWNIIEVPPNQHVPKPFYLLLEPRACGTCPPDYVEINGEKYCGE 400
 Qy 241 RSQFVTSNKNKTVTRHSHQSTYTDGFLAEYSYSDSDPCPGQFTTRGCRKLRCD 300
 Db 401 KEFVVSNNSSKMSVRFVSDSYTDGFLAEYSYSDSDPCPGQFTTRGCRKLRCD 460
 Qy 301 GMADCTHSDRLNCSGAGHGFCKN-KFCKPLFWVCDVNDGSDNDEGCGSCPAOTFR 359
 Db 461 GMADCTHSDRLNCSGAGHGFCKN-KFCKPLFWVCDVNDGSDNDEGCGSCPAOTFR 519
 Qy 360 CSNGKLSKSGQCGKDDCGDGSDBASCPRVNVYCTKTYRCLNGICLSKGNPCDCKE 419
 Db 520 CSNGKLSKSGQCGKDDCGDGSDBASCPRVNVYCTKTYRCLNGICLSKGNPCDCKE 579
 Qy 420 DCSYGSDEK--DCDCGLRSTRQARVVGCTDADGEGEMVQVSHALGQGHICGASLSRN 477
 Db 580 DCSYGSDEK--DCDCGLRSTRQARVVGCTDADGEGEMVQVSHALGQGHICGASLSRN 639
 Qy 478 MLVSAACHYIDDRGFRSDPTQMTAFGLHDOSORSAPOVERBLKRIISHPFENDPTFD 537
 Db 640 MLVSAACHYIDDRGFRSDPTQMTAFGLHDOSORSAPOVERBLKRIISHPFENDPTFD 699
 Qy 538 YDIALLELEKPAEYSSWVRPCLPDASHVPPAGKAIWVTMGHTQYGTGALILQKGEIR 597
 Db 700 YDIALLELEKPAEYSSWVRPCLPDASHVPPAGKAIWVTMGHTQYGTGALILQKGEIR 759
 Qy 598 VINOTTCENILPQOITRRMNCVGFSLSGVDSCGDSGGLSYVADGRITQAGVSVSGDG 657
 Db 760 VINOTTCENILPQOITRRMNCVGFSLSGVDSCGDSGGLSYVADGRITQAGVSVSGDG 819
 Qy 658 CAORNKPGVYTRLPDPDWIKENTGV 683
 Db 820 CAORNKPGVYTRLPDPDWIKENTGV 845

RESULT 7
 Q9DGR1 PRELIMINARY; PRT; 845 AA.
 AC Q9DGR1;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
 DE Homology of human MT-Spl.
 GN Name=XMT-Spl;
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
 OC Xenopodidae; Xenopus.
 CX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20363741; PubMed=10903452;
 RA Yamada K., Takabatake T., Takeshima K.,
 RT "Isolation and characterization of three novel serine protease genes
 RT from Xenopus laevis."
 RL Gene 252:209-216(2000).
 CC -1- SIMILARITY: Belongs to peptidase family S1.
 DR EMBL: AB038498; BAB08218.1; -.
 DR HSSP: P00760; 3BTH.
 DR MEROPS; S01.050; -.

Query Match 60.2%; Score 2280.5; DB 2; Length 845;
 Best Local Similarity 57.3%; Pred. No. 5.3e-152;
 Matches 393; Conservative 125; Mismatches 161; Indels 7; Gaps 4;

1 MAERVVMLPFRASLKSFFVTSVVAFPTDSKTVQRTQDNCSPGLHARGVLEWRFTTPG 60

DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000859; CUB.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR InterPro; IPR009003; Pept_Ser_Cys.
DR Pfam; PF00431; CUB; 2.
DR Pfam; PF00057; Ldl_recept_a; 4.
DR Pfam; PF00069; Trypsin_1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00261; LDLRECEPTOR.
DR SMART; SM00042; CUB; 2.
DR SMART; SM00192; LDla; 4.
DR SMART; SM00020; TRYP_SPC; 1.
DR PROSITE; PS01180; CUB; 2.
DR PROSITE; PS01209; LDla_1; 2.
DR PROSITE; PS50068; LDla_2; 4.
DR PROSITE; PS50240; TRYPIN_DOM; 1.
DR PROSITE; PS00134; TRYPIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPIN_SER; 1.
DR Hydrobase; Protease; Serine protease.
SQ SEQUENCE 845 AA; 93597 MW; 7FD7E62851A758B7 CRC64;

Query Match 59.5%; Score 2254.5; DB 2; Length 845;
Best Local Similarity 57.0%; Pred. No. 3.6e-150;
Matches 391; Conservative 123; Mismatches 165; Indels 7; Gaps 4;

QY 1 MAERVVMLPPRRASLSFVTVSVAPPTDSKYQVRQDNDSGSGHARVGLMRFTTPG 60
DB 164 ISELKLPVNPOR---TFALDSIVAYPTDQLARVFNKSCAYFLHSSNGVAKFSSFG 220
QY 61 FPDSPYPAHARCOVALRGDSDVSLTFRSPDLASCERGSDELVTYNTLSPEMPLAVQ 120
DB 221 FPDSPYPAHARCOVALRGDSDVSLTFRSPDLASCERGSDELVTYNTLSPEMPLAVQ 120
QY 121 LCGTPSPYNLTFRHSQNVLLITLITNTERHHPGFATFQLPMSGGRRLKRAQGTN 180
DB 281 LCGIYPSYNLTFRHSQNVLLITLITNTERHHPGFATFQLPMSGGRRLKRAQGTN 180
QY 181 SPYRPGHYPRINDCTNMIETVNNHNVKRFKFLLEPRACGCPDYVEINEKTCGE 240
DB 341 SPYRPGHYPRINDCTNMIETVNNHNVKRFKFLLEPRACGCPDYVEINEKTCGE 240
QY 241 RSQFVTVSNNSKIVRHSQSYDTGTFLAEYLSYSDSDPCPGQFTCRGACIKELRCD 300
DB 401 KEFVTVSNNSKIVRHSQSYDTGTGTFLAEYLSYSEPRNFCPDQFTCRGACIKELRCD 300
QY 301 GMADCTDHSDELNCSGAGHQTCKN-KFCKPLFWCDVNDGSDNSDEQSCCPAQTFR 359
DB 461 GMADCEBFSDEMSTCTA-LQFRCVNSKLCRPSYFICDGVNDGSDNSDELAKCPCNNTPK 519
QY 360 CSNCKCLSKSQGCKGKDDGSDGSDASCPKYNVTCTKHRYRCNLGCLSKNGECCKGE 419
DB 520 CGNCKCLPSQCKCRVNCGSDGSDASCPKYNVTCTKHRYRCNLGCLSKNGECCKGE 419
QY 420 DCSGDSDEK--DCDGLRSTFRQAVVGTADAGEPMQVSLALQGHICGASLSPN 477
DB 580 DCSGDSDEK--DCDGLRSTFRQAVVGTADAGEPMQVSLALQGHICGASLSPN 477
QY 478 MLVSAHCYIDRGRFRYSDPTQWTAFLGLHDQSGRASPAGVQERRKRIISHPFNDFTFD 537
DB 640 MLVSAHCYIDRGRFRYSDPTQWTAFLGLHDQSGRASPAGVQERRKRIISHPFNDFTFD 537
QY 538 YDIALLELRPAEISSWVRPCLPDASHVPVAGAAIWTGKHTQYQYGTGALLIKKEIR 597
DB 700 NDIAVLELRPAEISSWVRPCLPDASHVPVAGAAIWTGKHTQYQYGTGALLIKKEIR 597
QY 598 VINQTEENLLPQVITPRMNCVGLSGVDSCQSDSGAPLSVADGRIFQAGVSWGSDG 657
DB 760 IINQTEENLLPQVITPRMNCVGLSGVDSCQSDSGAPLSVADGRIFQAGVSWGSDG 657

QY 658 CAQRNKGYYTRLPFRDMIKENTGV 683
DB 820 CARRNKGYYTRLPFRDMIKENTGV 845

RESULT 8

Q6DEV0 PRELIMINARY; PRT; 663 AA.
AC Q6DEV0; (TEMBLrel. 28, Created)
DT 01-OCT-2004 (TEMBLrel. 28, Last sequence update)
DE Hypothetical protein.
OS Xenopus tropicalis (Western clawed frog) (Silurana tropicalis).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodidae; Xenopus.
OX NCBI_TaxID=8364;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=2238257; PubMed=12477932;

RA Strausberg R.L., Beilgold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Stammen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusik A.K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udell T.B., Toshitoki S., Carninci P., Prange C.,
RA Raha S.S., Loggellano N.A., Peters G.J., Abramson R.D., Mullaney S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzyzanski M.I., Skalska U., Smalls D.E., Schmechel A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA Klein S., Strausberg R.,
RI Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC076994; AAH76994.1; -.
KW Hypothetical protein.
SQ SEQUENCE 663 AA; 73914 MW; 46B2A56C657C7739 CRC64;

Query Match 40.7%; Score 1543; DB 2; Length 663;
Best Local Similarity 41.5%; Pred. No. 3.3e-100;
Matches 278; Conservative 91; Mismatches 115; Indels 186; Gaps 4;

QY 17 KSFVTVSVAAFPDSTVQRTQDNDSGSGHARVGLMRFTTPGPDSPYPAHARCOVAL 76
DB 177 RAFAVSLVAYPPDIPDIAINFKNSSCAFLHSSAGVMTGSSFGPDPTPPVPAHARCOVAL 236
QY 77 RGADADVLSLTFPSFPLASGDERGSDLVYNTLSPEMPLAVQLGTPSPYNLTFRHS 136
DB 237 RADAGMIRLKEFTFMEKCKANAGFVWYVDSLSIEPRAQRLCGIYPSYNLTFRHS 236
QY 137 QNVLTLLITNTERHHPGFATFQLPMSGGRRLKRAQGTENSPPYRPGHYPRINDCTW 196
DB 297 SNVMTLVITVNDNGKFPGLAEISQFPKTSICGRIYRDSGVTSYSPFGHYPKIECIW 356
QY 197 NIEVPNNQHVAKRFEKPYLLEPRACGCPKDYVEINEKTCGESEQFVTVSNNSKIVR 256
DB 357 DIQVPNNKFEVKAIRFNFFYLAEPVPTKCTKQFVEINGQYCCERKFPVTVSNNSKIVR 416
QY 257 FHDSDYTDGTFLAEYLSYSDSDPCPGQFTCRGACIKELRCDGMADCTDHSDELNCSG 316

```

Db 417 FVSDQSYTDGTGTAETLYSYEPNRPDPQFACKSGRCIRLIDQKCDGMNDCEDPSDEKSTC 476
Qy 317 DAGHQFCNK-KFKCPPLFWVCDSVNDGSDNSDEGSCSPACTPRCSNGKCLSKSQCCNGK 375
Db 477 TA-LQFCTNSKLTCPSTFVCDGVNDCSDSDBLACQCPNNYTCGNGKCLPSQKCDRT 535
Qy 376 DDCDGDSDASCPKRVNVTCTKTYRCINGLCLSKGNPECDGKEDSDSDSDE--KDCDCG 433
Db 536 DNGDGDSDASCPKRVNVTCTKTYRCINGLCLSKGNPECDGKEDSDSDSDSDE--KDCDCG 595
Qy 434 LRSTQARVVGCTDADGEMWPMQVSLHALQCGHI CGASLISPMWLYSAHCHYIDRGR 493
Db 596 KRPFCKSRIVGVNA----- 611
Qy 494 YSDPTQWTAFLGLHDQSORASPGVERLKRILSHPFNDPTFPYDIALLEKPAEYSS 553
Db 612 ----- 611
Qy 554 MVRPCLPDASHVPAGKAIWVTGMHTQYGTGALILQKGEIRVINGTCENLLPOQIT 613
Db 612 ----- 611
Qy 614 PRMVCVGLSGVSDSCGDSGGLSVADGRIFQAGVNSGDCQAQRNKGVTTLPLF 673
Db 612 -----DSGGLSVSLNNKYLAVIGVSGECARNRKGVYTRVAMM 653
Qy 674 RDMKENTGV 683
Db 654 RDMTRDTGL 663

RESULT 9
Q2RTY8 PRELIMINARY; PRT; 572 AA.
AC Q2RTY8;
DT 01-MAR-2004 (TREMblrel. 26, Created)
DT 01-MAR-2004 (TREMblrel. 26, Last sequence update)
DT 01-OCT-2004 (TREMblrel. 28, Last annotation update)
DE Type II transmembrane serine protease 7 precursor (Hypoetical
DE protein FLJ10088).
GN Name=TMPRSS7;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22722134; PubMed=12838346;
RA Puente X.S., Sanchez L.M., Overall C.M., Lopez-Otin C.;
RT "Human and mouse proteases: a comparative genomic approach.";
RL Nat. Rev. Genet. 4:544-558(2003).
RN [2]
RP SEQUENCE FROM N.A.
RA Nimomiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
RA Futuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,
RA Katsuya N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,
RA Kakehori K., Takahashi-Fuji A., Oshima A., Sugiyama A., Kawakami B.,
RA Suzuki Y., Sugano S., Nagahari K., Masuo Y., Nagai K., Isegai T.;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Belongs to peptidase family S1.
EMBL: BN000125; CAG67577.1; -.
DB MEROPS; S01.072; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004235; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000859; CUB.
DR InterPro; IPR02172; LDL_receptor_A.

```

```

DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR Pfam; PF00431; CUB; 1.
DR Pfam; PF00057; LDL_recept_a; 2.
DR Pfam; PF00089; Trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00261; LDLRECEPTOR.
DR PROSITE; PS01180; CUB; 2.
DR PROSITE; PS01209; LDLRA_1; 1.
DR PROSITE; PS00068; LDLRA_2; 2.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolyase; Protease; Serine protease; Signal; Transmembrane.
FT SIGNAL 1 23
FT CHAIN 24 572 type II transmembrane serine protease 7.
SQ SEQUENCE 572 AA; 64041 MW; 0458CBEA96BA285 CRC64;

Query Match 26.9%; Score 1017.5; DB 2; Length 572;
Best Local Similarity 36.3%; Pred. No. 2.8e-63;
Matches 226; Conservative 98; Mismatches 223; Indels 75; Gaps 21;

Qy 72 CQWALRDADSVLSLTFRSFDLASCDERGSGLVTVNTLSPEMPEHALVOLCTTPSYNL 131
Db 2 CPEFLVAIVGLYLRISIKSIQI-EADNCVDSLSLYDSLPIRSHSLYRIC-EPTRTLM 58
Qy 132 TFSQSNVLLITLTITNERBHPGEATFPOLPHMSGGRLRK---AQTSPSPYRGH 187
Db 59 SFVSTNNMLVTFKSPHRIKLSIRAYFVYIPECKCENTVLVDITFEGKISSPYPSY 118
Qy 188 YPNPIDCTNANIEVNNQNVHVRKFFYLLEPRRACGTPKDYVEINKEKYCGE--RSQFV 245
Db 119 YPKCKCTWKFQRT-SLTSLGALKFNYSITKSKMCKEHWMEINHWCGSYMHDQTI 177
Qy 246 VTSNSKRTVRFHSDQSYDTGFLAETLYSDSDPCP-GQFTCTGTCIKELRCQGMAD 304
Db 178 FRVPSPLVHIQLQCSSRLSDKPLAEGSYNISQPCVGSFRSSGLCVPAQRCQGVND 237
Qy 305 CTHSDSLNCSQAGHQFTCKNFKC---KPLFWVCDSVNDGSDGSGSCPAQTRFCS 361
Db 238 CFDSDEL-----FCVSPDP-----ACNTSFR-Q 261
Qy 362 NGKCLASQOQNGKDDCGSDSDASCPKRVNVTCTKTYRCINGLCLSKGNPECDGKEDC 421
Db 262 HGRPLI-----CDGRDDENRDENGNC--TQSIPENNRTFPGANDICRKNKACDGVDC 314
Qy 422 SDGSDERDCGGLRSFTRQARVVGCTDADGEMWPMQVSLHALQCGHI CGASLISPMWLV 481
Db 315 PDGSDERDCGGLRSFTRQARVVGCTDADGEMWPMQVSLHALQCGHI CGASLISPMWLV 372
Qy 482 AARCYIDRGRFVSDFPQWTAFLGLHDQSOR--APVQERLRKRILSHPFNDPTFPDYD 539
Db 373 AARCF--HGNRLSDPTFPAHGLGMVQGNKAFVSP-----VRLVHRYTNSQTDYD 423
Qy 540 IALTELE--KPAEYSWVRPCLPDASHVPAGKAIWVTGMHT-QYGTGALILQKGEI 596
Db 424 IALLQLSIAPETIKQLQIGTICIPRQGRVRSSEKCMVITMGKRNHEDNNKGLVLAQAV 483
Qy 597 RVINGTCENLLPQOITPRMVCVGLSGVSDSCGDSGGLSVADGRIFQAGVNSGDCQAQRN 655
Db 484 ELIDQTLCTVSTY-GITSRMLCAGINSGKRDACKGSDGGLSGRKSQDKWIIILGIYSWG 542
Qy 656 DGCAGRNKPGVYTRLPFRDWI 677
Db 543 HGGGRNPFPGVYTRVSNFPMI 564

RESULT 10
ID BAD18401 PRELIMINARY; PRT; 572 AA.
AC BAD18401;
DT 12-MAY-2004 (TREMblrel. 27, Created)
DT 12-MAY-2004 (TREMblrel. 27, Last sequence update)

```

DT 12-MAY-2004 (TReMBLrel. 27, last annotation update)
 DE CNNA FLJ16088 f1s, clone NT2RP7008435, weakly similar to EPITIMIN (EC 3.4.21.-).
 OS Homo sapiens (Human).
 OC Bakaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Euteheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Nishimura K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
 RA Furuya T., Takahashi M., Kikawa E., Omura Y., Abe K., Kamihara K.,
 RA Katsuta N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,
 RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
 RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
 RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,
 RA Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
 RA Suzuki Y., Sugano S., Nagahara K., Masuho Y., Nagai K., Isegai T.,
 RT "NEDO human cDNA sequencing project."
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AKJ1211; BAD18401.1; -
 SQ SEQUENCE 572 AA; 64041 MW; 0458CEA996EA285 CRC64;

Query Match 26.9%; Score 1017.5; DB 2; Length 572;
 Best Local Similarity 36.3%; Pred. No. 2.8e-63;
 Matches 226; Conservative 98; Mismatches 223; Indels 75; Gaps 21;

QY 72 CQNALRDADSVLSLTPRFDLASCDEGSDLVTVVNTLSPMEMPHALVOLCGTYPPEYNL 131
 DB 2 CHTFLVAIVGLVRLSLKSIQI-BADNCVDSLTIVSLPIRSILYRIC-EPTITLM 58
 QY 132 TFFSSQVLLITLTTERHPGFEATPQLPRMSCGGRLRK--AQGTSPYYPGH 187
 DB 59 SFVSTNMMLVTFKSPHRIIRLSGIRAFEVLPBQKCNVTLVKDITFEGKISSPYPSY 118
 QY 188 YPPIDCTANIEVNNHNVKRFELBRACGCPDYVINGEKYGE--RSQFV 245
 DB 119 YPPCKCKTMRQGT-SLSTLGLALFVNYSTYKSKCEHGHWINSHMTGSGYMDHOTI 177
 QY 246 VTSNSNKTIVRFHSDQSYDTGFLAEVLSYDSDPCE-GQFTCTGRCIRKELRCDGMAD 304
 DB 178 FRVPSPLVHQLQSSSLSDKPLAEGSYNISQPCVSVGRSSGLCVQAQRCDVDND 237
 QY 305 CTDHSDFLNCSDAHQFTCKNFKC--KPLFWVCDSVNDGNSDQSCSPAQTRFCS 361
 DB 238 CPDESDEL-----FCVSPQF-----ACNTSSPR-Q 261
 QY 362 NGKLSLSSQCGNGKDDGSDSCPKAVVCTKHTVCLNGLCISKNGPDCGKEDC 421
 DB 262 HGPLI-----CDGFRDCENGDEBQNC--TOSIPCNNTFCGNDICFRKQNAKCDGTVD 314
 QY 422 SDGSDKDCDQGLRSTFRQARVVGTDADGEMFQVSLHALCGGHTCGASLSPNVLVS 481
 DB 315 PDGSDGEGCTCS-RSSSALHRIIGDTTLBEGMWQVSLHFGVAGY-CGASVISREMLLS 372
 QY 482 AAHCYIDRGRFRSDPQWTAFLGLHDQSQRS--APGVQERRLKRIISHPPENDFTFYD 539
 DB 373 AAHCF--HGRRLSDPRTWTAHLCMYVQGNAKFVSP-----VRIYVHEVNSQTFDYD 423
 QY 540 IALLLEF--KPAEYSSWVRPTCLPDASHVPRAGAIWVTGNGHT-QVCGGALILQGEI 596
 DB 424 IALLQLSLAMPETIKOLIPICIPPTGQVRYSGKCVWTVGRRHEDNKGSLVLQQAELV 483
 QY 597 RVINQTTCCENLLPQOITPRMNCVFLSGVDS CGDSDGGLPS--SVEADGRIFQAGVSWG 655
 DB 484 ELIDQTLCSVTY-GIITSRLMCGAGINSGKDACGDSGGPLSCRSKSDGKMLILGIYSWG 542
 QY 656 DGCQRNKKPGVYTRLPFRDVI 677
 DB 543 HGGCRPMFPGVYTVFSNFWPI 564

ID QGBIK6 PRELIMINARY; PRT; 572 AA.
 AC QGBIK6;
 DT 01-MAR-2003 (TReMBLrel. 23, Created)
 DT 01-MAR-2003 (TReMBLrel. 23, last sequence update)
 DT 01-MAR-2004 (TReMBLrel. 26, last annotation update)
 DE Mus musculus adult male corpora quadrigemina cDNA, RIKEN full-length
 DE enriched library, clone:B230219123 product:weakly similar to BLOOD
 DE COAGULATION FACTOR XI.
 GN Name=B230219123Rik;
 OS Mus musculus (Mouse).
 OC Bakaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Euteheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;
 RC MEDLINE=9279253; PubMed=10349636;
 RA Carninci P.; Hayaehizaki Y.;
 RT "High-efficiency full-length cDNA cloning."
 RL Meth. Enzymol. 303:19-44(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;
 RC MEDLINE=21085660; PubMed=11217851;
 RA RIKEN FANTOM Consortium;
 RT "Functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:685-690(2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;
 RC THE FANTOM Consortium;
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs."
 RL Nature 420:563-573(2002).
 RN [4]
 RP SEQUENCE FROM N.A.
 RA STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;
 RC MEDLINE=20499374; PubMed=11042159;
 RA Carninci P.; Shibata Y.; Hayatsu N.; Sugahara Y.; Shibata K.; Itoh M.,
 RA Konno H., Okazaki Y., Muramatsu M., Hayaehizaki Y.;
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to
 RT prepare full-length cDNA libraries for rapid discovery of new genes."
 RL Genome Res. 10:1617-1630(2000).
 RN [5]
 RP SEQUENCE FROM N.A.
 RA STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;
 RC MEDLINE=20530913; PubMed=11076861;
 RA Shibata K.; Itoh M.; Aizawa K.; Nagaoka S.; Sasaki N.; Carninci P.,
 RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H.; Itoh M.,
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishino T., Harada A.,
 RA Yamamoto R., Matsunoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
 RA Fujikawa S., Inoue K., Togawa Y., Izawa M., Ohara E., Wachihi M.,
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayaehizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format
 RT sequencing pipeline with 384 multicapillary sequencer."
 RL Genome Res. 10:1757-1771(2000).
 RN [6]
 RP SEQUENCE FROM N.A.
 RA STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;
 RC Aachli U.; Aizawa K.; Akimura T.; Arakawa T.; Bono H.; Carninci P.,
 RA Fukuda S.; Furuno M.; Hanagaki T.; Hara A.; Hashizume W.,
 RA Hayaehizaki K.; Hayatsu N.; Hiramoto K.; Hiraoka T.; Kasukawa T.,
 RA Hori F.; Imclari K.; Ishii Y.; Itoh M.; Kagawa I.; Kasukawa T.,
 RA Katon H.; Kawai J.; Kojima Y.; Kondo S.; Konno H.; Kouda M.; Koya S.,
 RA Kurihara C.; Matsuyama T.; Miyazaki A.; Murata M.; Nakamura M.,
 RA Nishi K.; Nomura K.; Numazaki R.; Ohno M.; Osakato N.; Okazaki Y.,
 RA Saito R.; Saitoh H.; Sakai C.; Sakai K.; Sakazume N.; Sano H.,
 RA Sasaki D.; Shibata K.; Shingawa A.; Shiraki T.; Sogabe Y.; Tagami M.,
 RA Tagawa A.; Takahashi F.; Takaku-Akchira S.; Takada Y.; Tanaka T.,
 RA Tomaru A.; Toya T.; Yasunishi A.; Muramatsu M.; Hayaehizaki Y.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.

RESULT 11
 QGBIK6

```

CC -1- SIMILARITY: Belongs to peptidase family S1.
DR EMBL; AK045663; BAC2448.1; -.
DR HSSP; P00760; IEZX.
DR MEROPS; S01.072; -.
DR MGD; MG1:2686594; B230219123R1K.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000859; CUB.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR InterPro; IPR009003; Pept_Ser_Cys.
DR Pfam; PF00431; CUB; 1.
DR Pfam; PF00057; LDL_recept_a; 3.
DR Pfam; PF00089; Trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00261; LDLRECEPTOR.
DR SMART; SM00042; LDLa; 1.
DR SMART; SM00192; LDLa; 3.
DR SMART; SM00020; TRYP_SPE; 1.
DR PROSITE; PS01180; CUB; 2.
DR PROSITE; PS01209; LDLRA_1; 1.
DR PROSITE; PS50068; LDLRA_2; 2.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR HydroLase; Protease; Serine protease.
KW SEQUENCE 572 AA; 63757 MW; 21C0EC523B5F4301 CRC64;

Query March 26.8%; Score 1015.5; DB 2; Length 572;
Best Local Similarity 36.3%; Pred. No. 3.9e-63;
Matches 225; Conservative 98; Mismatches 226; Indels 71; Gaps 21;

QY 72 CQNALRGDADSVLSLTFRSPDLASCDEKSGSDLVYVNTLSPEMHALVOLGTYPPSYNL 131
DB 2 CHPFLVAIVGLIRLSIESIQTL-EADNDITDLSLTYDLSLIRSAIIVRIC--EPTRTIM 58

QY 132 TFHSSQNVLLITLTNTERRRPGFBAITFQLPFRSSCCG-----RLRAQGTNSPYPG 186
DB 59 SFVSTNNMLVILKSPYRRLAGIRAYFEVIPR-OKCESTLIVKINSFEKISSPYPS 117

QY 187 HYPERNDICTNIEVYNNQHVAVRFPFYLEPRACGCPDYVINEKXCGE--RSQF 244
DB 118 YVPRCKCKTWTFFQTL-SLSTGLALKFTYVSTTKSAKCEHWEINEMTCGSIYMDHET 176

QY 245 VVTSNSNKITVRFHSDSYDTGTFLAELYSDSDPCP-GQFTGRTGRCIRKEILRCGWA 303
DB 177 IFRPSPPLVHQLQCSSRLSKPLVEYGVNISOQCPAGSPRCSGGLCTVPAQRCQGVN 236

QY 304 DCTHSDDELNCSAAGHOFTCKNFKCKPLFVVCDSVNDGNSDQSGSCPAQTPRCGNG 363
DB 237 DCFPESDELFCVT-----VKP-----ACNSSFR-OHG 263

QY 364 KCLSKSQQCNKDDCGSGDEASCPKVVVVTCTKTYRCLNGCLSKNPECDKEDPSD 423
DB 264 PLV-----CDGFRCEBDQDEQNCNR--SLPCTSRTEPCGNDICRKNACCDGIVDPD 316

QY 424 GSDKDCDCGLRSTFRQARVVGCTDADGEMPMQVSLHALGQGHICGASLSPMLVSA 483
DB 317 GSDDEGGCS--RSSSPFLHRIVIGSDSDGTPPMQVSLHFSVSA--CGASVLSRMLSSA 374

QY 484 HCYIDDRGFYSPTQMTAFGLHDQGRS--APGVORRLKRIISHPFNDFPFYDIA 541
DB 375 HCF---HGNRLSDPTPTAHLGMYQGNAKFISP-----VRAIVHEVYNSQTFDYDIA 425

QY 542 LLELE--KPAVSSAVVRPCLPDASHVPAKAIIVMGWHT-QYCGGALILKGRIRV 598
DB 426 LLDGSIAMPELTKQIDPICIPLPGQKVRSGKCVTGMGRHEDSGSVLQDAEVEL 485

QY 599 INQTTCEMLPQOITPRMVCVGLSGVDSCGDSGGPLS--SVEADGRIPQAGVNSWDG 657
DB 486 IDQTVVAVSTY-GITTSRLCAGVWVGSKSDACKGDSGGPLSRARRSDGMWLTGIVSWHG 544

```

```

QY 658 CAQRNKPQYTRLPJLFRDWI 677
DB 545 CGRNPFGVYTRVSSFPVWI 564

RESULT 12
ID Q6PF94 PRELIMINARY; PRT; 799 AA.
AC Q6PF94
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
OS Trypsin protein.
DS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N. A.
RC STRAIN=FVB/N; TISSUE=liver;
RX MEDLINE=22388257; Pubmed=12477932;
RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares W.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udell T.B., Toshlyuk S., Caraminci P., Prange C.,
RA Rana S.S., Loughellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Boeak S.A., McGwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettleman W., Madan A., Rodriguez S., Sanchez A.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butcherfield Y.S.,
RA Krzyzanski M.I., Skaleka U., Smalins D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Maitra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N. A.
RC STRAIN=FVB/N; TISSUE=liver;
RA Strauberg R.;
RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Belongs to peptidase family S1.
DR EMBL; BC057674; AAH57674.1; -.
DR GO; GO:0005886; C:plasma membrane; IDA.
DR InterPro; IPR000859; CUB.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR InterPro; IPR009003; Pept_Ser_Cys.
DR Pfam; PF00057; LDL_recept_a; 3.
DR Pfam; PF00089; Trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00261; LDLRECEPTOR.
DR SMART; SM00042; TRYP_SPE; 1.
DR SMART; SM00192; LDLa; 3.
DR PROSITE; PS01180; CUB; 1.
DR PROSITE; PS01209; LDLRA_1; 1.
DR PROSITE; PS50068; LDLRA_2; 3.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR HydroLase; Protease; Serine protease.
KW SEQUENCE 799 AA; 89557 MW; 16315A646AD5288 CRC64;

Query March 26.8%; Score 1014; DB 2; Length 799;
Best Local Similarity 34.4%; Pred. No. 7.2e-63;
Matches 213; Conservative 107; Mismatches 233; Indels 66; Gaps 15;

```

```

QY 72 COMALRGDADSVSLTFRSFDLASCDERGSDLVTYVNTLSPMEPHALVOLCGTYPSPSYNL 131
D 233 CLMHLQGPEDLMITKRL-EMTRVDCRDR---VAMTDAAGPLERKRLITSYVGSRGQRPVM 287
QY 132 TFHSSQVNLITLITNTERRHNGF---EATFQOLPRMS-SCGRLRKAQGTENSPYYPG 186
D 288 EVLASGSVMAVVMKKGMHSYDPLFLSVKSAFODCVNLTLEGR-L-DTQGFRLTPYPS 346
QY 187 HYPENIDCTNIEVPPNNQHVKRFKFFYLEPRRACGTCPDYVEINGEKYCGER----- 241
D 347 YTSSTHCSMHLTYPSLDYGLALMFDAYALRRQKYNRLCTQGGMTQNRRLCGFRLTQPY 406
QY 242 SQFVVTNSNKTIVRFHSDSYDTDTGFLAEYLSYDSSDPGQFTCRGCRIRKELRCDG 301
D 407 AERIPWASDGVNTINTSQTSLTGPVQVYSLYNSQDPCRGELGCV----- 454
QY 302 WADCTHSDSLNCSCDAGHQTCKNFKCPLFWVCDVNDGSDSEQSCSPAQTFR- 360
D 455 -----NGLCVP---ACDGIKCCPNGLDRNCVCRAM-FOCQ 486
QY 361 SNGKCLSKSQCKGKDDCGSDGSDASCPKYNVYCTKHTRYCLNGCLSKSGNPECCKED 420
D 487 EDSTCISLPRVCDRQDPCLNGSDEQCE--GVPCGFTTQCEPDRSCVCKRPNPECQSD 544
QY 421 CSDGSDKDCDGLRSTFRQARVVGTDADGEMPMQVSLHALQGHICGASLISPMVLV 480
D 545 CRDGSDEHCDGCLQGLS--SRIVGTVSSEGEPMQASLQIRGR-HICGALILADRMI 601
QY 481 SAACHYIDDRGFRYSDPTQWTAFLGLHDQSRAPGVQERRLKRIISHPFNDFTDYDI 540
D 602 TAAHCFQED---SMASPKLMTVFLGKMRQNSR--WPEVSFKVSLFLPHYHEEDSHDYD 657
QY 541 ALLELEKPAEYSMWVRPCLPDASHVFPAGKAIWVTGMGHTQVGTGALILQKEIRVIN 600
D 658 ALLQDLHPVYSAVVRVCLPARSHFEPQHCWITGKAGRGSPVNTLQKDVQVLP 717
QY 601 QTTCEMLLPQOITPRMNCVGLSGVSDSCGSDGSPLSVEADGRIFQAGVSMGDCAQ 660
D 718 QDLCSEAYRYQVSPRMLCAGYRKGKDAQCQDSGSPVLCRBPGRWFLAGLVSMGLCGR 777
QY 661 RNKRGVYTRLPFRDWKE 679
D 778 PNFFGYVTRVTRVINMIQ 796

```

RESULT 13
AAH57674 PRELIMINARY; PRT; 799 AA.

AC AAH57674;
DT 02-MAR-2004 (TREMBlrel. 27, Created)
DT 02-MAR-2004 (TREMBlrel. 27, Last sequence update)
DE 02-MAR-2004 (TREMBlrel. 27, Last annotation update)
OS Mus musculus (Mouse).
OC Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=liver;
RX MEDLINE=23388257; PubMed=12477932;
RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Bueltow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heien F.,
RA Dlatcenko L., Maruska K., Farmer A.A., Rubin G.M., Hong L.,
RA Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uedin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.U., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

```

RA Fahy J., Helton E., Kesteman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchan J.W., Green R.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzyzanski M.I., Skalka U., Smalins D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Maiz M.A.;
RT "generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=liver;
RA Strauberg R.;
RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC057674; AAH57674.1; -.
SQ SEQUENCE 799 AA; 89557 MW; 16315A646A4D5288 CRC64;

Query Match 26.8%; Score 1014; DB 2; Length 799;
Best Local Similarity 34.4%; Pred. No. 7.2e-63;
Matches 213; Conservative 107; Mismatches 233; Indels 66; Gaps 15;

QY 72 COMALRGDADSVSLTFRSFDLASCDERGSDLVTYVNTLSPMEPHALVOLCGTYPSPSYNL 131
D 233 CLMHLQGPEDLMITKRL-EMTRVDCRDR---VAMTDAAGPLERKRLITSYVGSRGQRPVM 287
QY 132 TFHSSQVNLITLITNTERRHNGF---EATFQOLPRMS-SCGRLRKAQGTENSPYYPG 186
D 288 EVLASGSVMAVVMKKGMHSYDPLFLSVKSAFODCVNLTLEGR-L-DTQGFRLTPYPS 346
QY 187 HYPENIDCTNIEVPPNNQHVKRFKFFYLEPRRACGTCPDYVEINGEKYCGER----- 241
D 347 YTSSTHCSMHLTYPSLDYGLALMFDAYALRRQKYNRLCTQGGMTQNRRLCGFRLTQPY 406
QY 242 SQFVVTNSNKTIVRFHSDSYDTDTGFLAEYLSYDSSDPGQFTCRGCRIRKELRCDG 301
D 407 AERIPWASDGVNTINTSQTSLTGPVQVYSLYNSQDPCRGELGCV----- 454
QY 302 WADCTHSDSLNCSCDAGHQTCKNFKCPLFWVCDVNDGSDSEQSCSPAQTFR- 360
D 455 -----NGLCVP---ACDGIKCCPNGLDRNCVCRAM-FOCQ 486
QY 361 SNGKCLSKSQCKGKDDCGSDGSDASCPKYNVYCTKHTRYCLNGCLSKSGNPECCKED 420
D 487 EDSTCISLPRVCDRQDPCLNGSDEQCE--GVPCGFTTQCEPDRSCVCKRPNPECQSD 544
QY 421 CSDGSDKDCDGLRSTFRQARVVGTDADGEMPMQVSLHALQGHICGASLISPMVLV 480
D 545 CRDGSDEHCDGCLQGLS--SRIVGTVSSEGEPMQASLQIRGR-HICGALILADRMI 601
QY 481 SAACHYIDDRGFRYSDPTQWTAFLGLHDQSRAPGVQERRLKRIISHPFNDFTDYDI 540
D 602 TAAHCFQED---SMASPKLMTVFLGKMRQNSR--WPEVSFKVSLFLPHYHEEDSHDYD 657
QY 541 ALLELEKPAEYSMWVRPCLPDASHVFPAGKAIWVTGMGHTQVGTGALILQKEIRVIN 600
D 658 ALLQDLHPVYSAVVRVCLPARSHFEPQHCWITGKAGRGSPVNTLQKDVQVLP 717
QY 601 QTTCEMLLPQOITPRMNCVGLSGVSDSCGSDGSPLSVEADGRIFQAGVSMGDCAQ 660
D 718 QDLCSEAYRYQVSPRMLCAGYRKGKDAQCQDSGSPVLCRBPGRWFLAGLVSMGLCGR 777
QY 661 RNKRGVYTRLPFRDWKE 679
D 778 PNFFGYVTRVTRVINMIQ 796

```

RESULT 14
TMS6 MOUSE STANDARD; PRT; 811 AA.

AC TMS6 MOUSE
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)

DE Transmembrane protease, serine 6 (EC 3.4.21.-) (Matrilysin-2).
 GN Name=tmpr96;.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
 RC STRAIN=C57BL/6J;
 RX MEDLINE=22755753; PubMed=12744720; DOI=10.1042/BJ20030390;
 RA Hooper J.D., Campagnolo L., Goodarzi G., Truong T.N., Stuhlmann H.,
 RA Quigley J.P.;
 RT "Mouse matrilysin-2: identification, characterization and comparative
 RT mRNA expression analysis with mouse hepsin in adult and embryonic
 RT tissues";
 RL Biochem. J. 373:689-702 (2003).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Liver;
 RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
 RA Nkaiato I., Oseato N., Saito R., Suzuki H., Yamataka I., Kiyosawa H.,
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gotohori T.,
 RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
 RA Schiraldi L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
 RA Blake U.A., Bradt D., Brusic V., Chochia C., Corbani L.E., Cousins S.,
 RA Datta E., Dragan T.A., Fletcher C.F., Forrest A., Frazer K.S.,
 RA Gaasterland T., Gariboldi M., Giesi C., Godzik A., Gough J.,
 RA Giamoni S., Gutentuch S., Hirokawa N., Jackson I.J., Jarvis E.D.,
 RA Kanai A., Kawai H., Kawasawa Y., Kedzierski R.M., King B.L.,
 RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
 RA Nagashima T., Numa K., Okido T., Pavan W.J., Petia G., Pessio G.,
 RA Petrovsky N., Pillai R., Pontius J.V., Qi D., Ramachandran S.,
 RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
 RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
 RA Vercato R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
 RA Wilming L.G., Wysshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
 RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
 RA Hitotsume-Kishikawa T., Kono H., Nakamura M., Sakazume N., Sato K.,
 RA Shiraki T., Waki K., Kawai K., Aizawa K., Arakawa T., Fukuda S.,
 RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
 RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shingawa A.,
 RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
 RA Birney E., Hayashizaki Y.;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs";
 RL Nature 420:563-573 (2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Blac N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Datsenko L., Marusik K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stetson M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ueda T.B., Toshlyuk S., Carninci P., Prange C.,
 RA Rata S.S., Loggellano N.A., Peters G.J., Abramson R.D., Mulhaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunnarac P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
 RA Villalón D.K., Wuzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Keating M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska J., Smallie D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).

RN [4]
 RP REVIEW.
 RX MEDLINE=22668120; PubMed=12784999;
 RA Netzel-Arnett S., Hooper J.D., Szabo R., Madison E.L., Quigley J.P.,
 RA Bugge T.H., Antalis T.M.;
 RT "Membrane anchored serine proteases: a rapidly expanding group of cell
 RT surface proteolytic enzymes with potential roles in cancer.";
 RL Cancer Metastasis Rev. 22:237-258 (2003).
 CC -1- FUNCTION: May play a specialized role in matrix remodeling
 CC processes in liver (By similarity).
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein.
 CC -1- TISSUE SPECIFICITY: Expressed at highest levels in liver, kidney
 CC and uterus.
 CC -1- DEVELOPMENTAL STAGE: Expressed at higher levels from 12.5 dpc to
 CC 15.5 dpc with a peak at 13.5 dpc. Expression in the developing
 CC liver as well as a restricted set of embryonic epithelial cells of
 CC the nasal cavity and pharyngo-lymphatic tubes.
 CC -1- SIMILARITY: Belongs to peptidase family S1.
 CC -1- SIMILARITY: Contains 2 CUB domains.
 CC -1- SIMILARITY: Contains 3 LDL-receptor class A domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AY240929; AAP69827.1; -;
 CC EMBL: AK004939; BAB23684.2; -;
 CC EMBL: BC029645; AAH29645.2; -;
 CC HSSP: P00760; IEXX.
 CC MEROPS: S01.308; -;
 CC MGD: MGI:1919003; Tmp96.
 CC GO: GO:0016021; C:integral to membrane; ISS.
 CC GO: GO:0005886; C:plasma membrane; IDA.
 CC GO: GO:0004252; F:serine-type endopeptidase activity; ISS.
 CC GO: GO:0001525; P:angiogenesis; ISS.
 CC GO: GO:0030198; P:extracellular matrix organization and bioge. . .; ISS.
 CC GO: GO:0042730; P:fibinolysis; ISS.
 CC GO: GO:0007242; P:intracellular signaling cascade; ISS.
 CC GO: GO:0006508; P:proteolysis and peptidolysis; ISS.
 CC InterPro: IPR000859; CUB.
 CC InterPro: IPR002172; LDL_receptor_A.
 CC InterPro: IPR001254; Peptidase_S1.
 CC InterPro: IPR001314; Peptidase_S1A.
 CC InterPro: IPR009003; Pept_Ser_Cys.
 CC Pfam: PF00057; Ldl_recept_a; 2.
 CC Pfam: PF00089; Trypsin_1.
 CC PRINTS: PR00722; CHYMOTRYPSIN.
 CC PRINTS: PR00261; LDLRECEPTOR.
 CC PROSITE: PS01180; CUB; 1.
 CC PROSITE: PS01209; LDLRA_1; 1.
 CC PROSITE: PS01209; LDLRA_2; 3.
 CC PROSITE: PS50246; TRYPSIN_DOM; 1.
 CC PROSITE: PS50246; TRYPSIN_HIS; 1.
 CC PROSITE: PS00133; TRYPSIN_SER; 1.
 KW Glycoprotein; Hydrolyse; Repeat; Serine protease; Signal-anchor;
 KW Transmembrane.
 FT DOMAIN 1 59 Cytoplasmic (Potential).
 FT TRANSMEM 60 80 Signal-anchor for type II membrane
 FT DOMAIN 81 811 protein (Potential).
 FT DOMAIN 213 336 Extracellular (Potential).
 FT DOMAIN 323 440 CUB 1.
 FT DOMAIN 445 477 CUB 2.
 FT DOMAIN 478 514 LDL-receptor class A 1.
 FT DOMAIN 518 555 LDL-receptor class A 2.
 FT DOMAIN 565 799 LDL-receptor class A 3.
 FT ACT_SITE 617 617 Serine protease.
 FT ACT_SITE 617 617 Charge relay system (By similarity).
 FT ACT_SITE 668 668 Charge relay system (By similarity).
 FT ACT_SITE 762 762 Charge relay system (By similarity).


```

FT CARBOHYD 138 138 N-linked (GlcNAc...) (Potential)
FT CARBOHYD 184 184 N-linked (GlcNAc...) (Potential)
FT CARBOHYD 216 216 N-linked (GlcNAc...) (Potential)
FT CARBOHYD 338 338 N-linked (GlcNAc...) (Potential)
FT CARBOHYD 433 433 N-linked (GlcNAc...) (Potential)
FT CARBOHYD 453 453 N-linked (GlcNAc...) (Potential)
FT CARBOHYD 518 518 N-linked (GlcNAc...) (Potential)
FT CONFLICT 690 690 P -> PP (in Ref. 2)
SQ SEQUENCE 811 AA; 90978 MW; 32EB3E7C3127801B CRC64;

Query Match 26.8%; Score 1014; DB 1; Length 811;
Best Local Similarity 34.4%; Pred. No. 7.3e-63;
Matches 213; Conservative 107; Mismatches 233; Indels 66; Gaps 15;

QY 72 COMALRDADSVSLTRFSPDLASCDGRSDLVVTYVTLSPMEPHALVQCGTYPSPYNL 131
Db CMLHLOEPEDMIVRL-EWTRVDCRR-----VAMYOAGLEGRLLTSYVCGSRQSPVM 299
QY 132 TFFSSQVLLTLTLTTRRRHGF-----EATFQLPMS-SCGRLKKAQGTENSPYPG 186
Db EVLASGSVMAVVMKKGMHSYDPLFLSVKSAVFQDCVNLTLEGRLL-DTQGLRTPYPS 358
QY 187 HYPNIDCTNNIEVPNNQHKVRFKFTYLLERACGTCPEDYVEINGEKTCGER----- 241
Db YSPSTHCSMHLTVPSLDYGLALMFDAYALRRQKYNRLCTQGWMIONRRLLCGFRLLQPY 418
QY 242 SDFVTSNKTIVRFHSDSYDTDTGLAEYLSYSDSPCEGQFTCTGRRCIRKELACDG 301
Db AERIPWASDGVITNFTSQTSLTGPQVYSLYNQDPCRGELCSV----- 466
QY 302 WADCTHSDINLSCDAHQFTCKNFKCKPLFWVCDSDVNDGSDSDQSCSPAQTRC- 360
Db 467 -----NGLCVP-----ACDGIKDCPNGLDERNCVCGRAM-FCQ 498
QY 361 SNGKCLSKSQCKNDKCGDSDASCPKNNVYCTHTYRCLNGLSLKGNPSCDKED 420
Db 499 EDSTCISLPRVCDBQPCPLNGSDDEQCE--GVPCGFTTQCEBRSCKKPNPCDQSD 556
QY 421 CSDSDKDCDCGSLRFTROARVVGTADBGEMVQVSLHALGQGHICGSLSPMWLV 480
Db 557 CRDSDQHCHDCGLQGS--SRIVGVSSBEGFPWQASLQIRK-HICGALIAADWVI 613
QY 481 SAACVYIDRGFRYSPTQWTAFLGLHDQGSAPAGVEERLKRILSHPFNFTFDYDI 540
Db 614 TAAACPFQED--SWASPKMTVFLGKRRQNSR-WPGEVSFVRSLLFLPHYHEDSHDYD 669
QY 541 ALLELEKPAEYSSNVRPICLPDASHVPAKAIWVTEGHTQYGTALLIQKEIVIN 600
Db 670 ALLQLDHPVYVSATVRVCLPARSHFEPGQHCWITGMAQREGGPPVSNLTQKVDVLP 729
QY 601 QTTCEMLTPOQITPRMNCVGFSLSGVSDSCGDSGPISSVEADGRIFQAGVSWDGCQAQ 660
Db 730 QDLCSEAYRYQVSRPMICAGYRRKKKDAACGDSGGLVCRBPSGRWFLAGLVSGLCGR 789
QY 661 RNKPGVYTRLPFLFDWIK 679
Db 790 PNFEGVYTRVTRVINVIOQ 808

RESULT 15
TMS6 HUMAN STANDARD; PRT; 811 AA.
AC O81UB0; O81UE2; O81XV8;
DT 10-OCT-2003 (Ref. 42, Created)
DT 05-JUL-2004 (Ref. 44, Last sequence update)
DE Transmembrane protease, serine 6 (EC 3.4.21-) (Matrilinease-2).
GN Name=TMPRSS6;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]

```

```

RP SEQUENCE FROM N.A. (ISOFORM 1).
RA Hooper J.D., Quigley J.P.;
RT "TMPRSS6, a new type II transmembrane serine protease.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBS databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20057165; PubMed=10591208; DOI=10.1038/990031;
RA Dunham I., Hunt A.R., Collins J.E., Bruskewich R., Beare D.M.,
RA Clamp M., Slink L.J., Alnsough R., Almeida J.P., Babbage A.K.,
RA Bagguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,
RA Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,
RA Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,
RA Clegg S.M., Cooley V.B., Cole C.G., Collier R.E., Connor R.,
RA Conway D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,
RA Dhami P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,
RA Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,
RA Gilbert J.G.R., Goward M.E., Grafham D.V., Griffiths M.N.D., Hall C.,
RA Hall R.E., Hall-Tamlyn G., Heathcote R.W., Ho S., Holmes S.,
RA Hunt S.E., Jones W.C., Kershaw J., Kimberley A.M., King A.,
RA Laird G.K., Langford C.F., Leversha M.A., Lloyd C., Lloyd D.M.,
RA Marlyn J.D., Mashreghi-Mohammadi M., Matthews L.H., Mccann O.T.,
RA McClay J., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J.,
RA O'Neill C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,
RA Phillips S.H., Plumb R.W., Ramsey H., Ranney Y., Rogers L., Ross M.T.,
RA Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,
RA Stender L.D., Spragon L., Steward C.A., Sulston J.E., Swann R.M.,
RA Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,
RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,
RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,
RA Wnoshina S., Kawasaki K., Sasaki T., Asakawa S., Kodoh J.,
RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyma S.,
RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,
RA Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,
RA Lewis J., Lewis S., Lin S.-P., Loh P., Malai E., Nguyen T., Pan H.,
RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shall S., Sloan D., Song L.,
RA Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,
RA Zhan M., Zhang G., Chisoe S., Murray J., Miller N., Mink P.,
RA Fulton R., Johnson D., Bemis G., Bentley D., Birdsall H., Bourne S.,
RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,
RA Hinds K., Kemp K., Latreille P., Layman D., Olesky P., Rohlfing T.,
RA Scheer P., Walker C., Wamsley A., Wohlmann P., Pepin K., Nelson J.,
RA Koef I., Bedell J.A., Hillier L.W., Mardis E., Waterston R.,
RA Wilson R., Emanuel B.S., Shaikh T., Kurashiki H., Salta S.,
RA Budarf M.L., McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E.,
RA Belman L., Kim U.J., Shizuya H., Simon M.I., Dumarski J.P.,
RA Peyrard M., Kedra D., Seroussi E., Franssen I., Tapia I., Bruder C.E.,
RA O'Brien K.P., Wilkinson P., Bodenteich A., Hartman K., Hu X.,
RA Khan A.S., Lane L., Tiliannu Y., Wright H.;
RT "The DNA sequence of human chromosome 22.";
RL Nature 402:489-495(1999).
RN [3]
RP SROUNCE FROM N.A. (ISOFORM 2).
RC TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Klausner R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schaller G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Schaefer J., Heich F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Brownstein M., Soares M.B., Bonaldo W.F., Casavant J.L., Scheetz T.E.,
RA Raba S.S., Loquellano N.A., Peters G.J., Abremson R.D., Mallary S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shewchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,
RA Schermer A., Schein U.E., Jones S.J.M., Maitra K.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";

```


Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).

RA [4] SEQUENCE OF 9-811 FROM N.A. (ISOFORM 1), AND TISSUE SPECIFICITY.

RP TISSUE=Fetal liver;

RC MEDLINE=22241917; PubMed=12149247; DOI=10.1074/jbc.M203007200;

RA Velasco G., Cal S., Quesada V., Sanchez L.M., Lopez-Otin C.;

RT "Matriptase-2, a membrane-bound mosaic serine proteasease predominantly expressed in human liver and showing degrading activity against extracellular matrix proteins";

RT J. Biol. Chem. 277:37637-37646 (2002).

RN [5] REVIEW.

RP MEDLINE=22668120; PubMed=12784999;

RA Netzel-Arnett S., Hooper J.D., Szabo R., Madison E.L., Quigley J.P., Bugge T.H., Antalis T.M.;

RT "Membrane anchored serine proteases: a rapidly expanding group of cell surface proteolytic enzymes with potential roles in cancer";

RT Cancer Metastasis Rev. 22:237-258 (2003).

CC -1- FUNCTION: May play a specialized role in matrix remodeling processes in liver.

CC -1- SUBCELLULAR LOCATION: Type II membrane protein.

CC -1- ALTERNATIVE PRODUCTS:

CC Event=Alternative splicing; Named Isoforms=2;

CC Name=1;

CC IsoId=Q8IU80-1; Sequence=Displayed;

CC Name=2;

CC IsoId=Q8IU80-2; Sequence=VSP_008379, VSP_008380;

CC Note=No experimental confirmation available;

CC -1- TISSUE SPECIFICITY: Liver specific.

CC -1- SIMILARITY: Belongs to peptidase family S1.

CC -1- SIMILARITY: Contains 2 CUB domains.

CC -1- SIMILARITY: Contains 3 LDL-receptor class A domains.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: AJ139876; CAC85953.1; ALT_INIT.

DR EMBL: AY055383; AA16413.1; -

DR EMBL: AY055384; AA16414.1; -

DR EMBL: AL022314; -; NOT_ANNOTATED_CDS.

DR EMBL: BC039082; AAH39082.1; -

DR HSSP: P00760; 1EXX.

DR Genew: HGNC:16517; TMPRSS6.

DR InterPro: IPR000859; CUB.

DR InterPro: IPR002172; LDL_receptor_A.

DR InterPro: IPR001254; Peptidase_S1.

DR InterPro: IPR001314; Peptidase_S1A.

DR InterPro: IPR009003; Pept_Ser_Cys.

DR Pfam: PF00431; CUB; 1.

DR Pfam: PF00057; LDL_recept_a; 2.

DR Pfam: PF00089; Trypsin_1.

DR PRINTS: PR00722; CHYMOTRYPSIN.

DR PRINTS: PR00261; LDLRECEPTOR.

DR SMART: SM00042; CUB; 1.

DR SMART: SM00192; LDLa; 3.

DR SMART: SM00020; Tryp_spc; 1.

DR PROSITE: PS01180; CUB; 1.

DR PROSITE: PS01209; LDLRA_1; 1.

DR PROSITE: PS00068; LDLRA_2; 3.

DR PROSITE: PS50240; TRYPSIN_DOM; 1.

DR PROSITE: PS00134; TRYPSIN_HIS; 1.

DR PROSITE: PS00135; TRYPSIN_SER; 1.

KM Alternative splicing; Glycoprotein; Hydrolase; Repeat;

KM Serine protease; Signal-anchor; Transmembrane.

FT DOMAIN 1 55 Cytoplasmic (Potential).

FT TRANSMEM 56 76 Signal-anchor for type II membrane protein (Potential).

FT DOMAIN 77 811 Extracellular (Potential).

FT DOMAIN 213 336 CUB 1.

FT DOMAIN 335 452 CUB 2.

FT DOMAIN 457 489 LDL-receptor class A 1.

FT DOMAIN 490 526 LDL-receptor class A 2.

FT DOMAIN 530 567 LDL-receptor class A 3.

FT DOMAIN 577 617 Serine protease.

FT ACT_SITE 617 617 Charge relay system (By similarity).

FT ACT_SITE 668 668 Charge relay system (By similarity).

FT ACT_SITE 762 762 Charge relay system (By similarity).

FT CARBOHYD 136 136 N-linked (GlcNAc...) (potential).

FT CARBOHYD 184 184 N-linked (GlcNAc...) (potential).

FT CARBOHYD 216 216 N-linked (GlcNAc...) (potential).

FT CARBOHYD 338 338 N-linked (GlcNAc...) (potential).

FT CARBOHYD 433 433 N-linked (GlcNAc...) (potential).

FT CARBOHYD 453 453 N-linked (GlcNAc...) (potential).

FT CARBOHYD 518 518 N-linked (GlcNAc...) (potential).

FT VARSPLIC 409 461 LCGRIQLQPAERIPVAVATGRTINFTSISLPGVRYH GLYNOSPPCGE -> YHPLSSIMLPLPPPPSPSTVTP SLEAQVNLRGAGASRGWMCQACCP (in isoform 2).

FT VARSPLIC /FTId=VSP_008379.

FT Missing (in isoform 2).

FT VARSPLIC /FTId=VSP_008380.

FT CONFLICT 116 116 A -> V (in Ref. 4).

FT SEQUENCE 811 AA; 8999 MM; 7BEF193P655DE9D CR64;

Query Match 26.6%; Score 1007; DB 1; Length 811;

Best Local Similarity 34.0%; Pred. No. 2,3e-62;

Matches 217; Conservative 99; Mismatches 229; Indels 94; Gaps 17;

67 PAH-ARCOMALRGADSVSLTFRSFDLASCDERGSGLTVYVNTTSPMEPHALVOL--C 122

238 PDHASSCLMHLQCPKDMKLRL-EWTLAECDR--LAMDVAGPLKRLITSYGC 292

123 GYPPSYNL------FHSSQNVLLTLTNTERRRHGEFATFQGLPRMSSCG 169

293 SRQPVVEVLASGAIMAVVWKKGLSYDPEFLSV-----QPVVFOACEVNLTL 341

170 GLRLKAGCTNSPPYRPHYEPNIDCTWNIENPNQHKVAFKFFYLLEPRRAGCTPKDY 229

342 DNRLDSSGVSLTPFPSPYSPQTHCSWMLTVPSLDYGLALMFAYALRRKXYDLPTQOQ 401

230 VEINGEKYCG-----ERSQFVTSNSNKKITVRFHSDSYTDYGLAEYLSYDSDPC 281

402 WTIQNRLLCGRLIQPAERIPVAVATG--ITINFSQSLSGPGRYHYGLYNSDPC 458

282 PGQFTCRTRGCIKRELCDGMADCTDHSDELNCSGAGHQTCKNFKCPLFWVCDVND 341

459 GGEFLGCV-----NGLCVF--ACDGVKD 479

342 CGDNSDEQCSCPAQTRFC-SNGKLSKSQCCNGKDDCGSDGASCPKXNVVTCRHTY 400

480 CPNGLDERNCVCRA-FTQCKEDSTCISLPRVCQGPCLNGSBEQOE--GVPCGTFP 536

401 RCLNGLCTSKGNPCDCKEDPCDSDKDCDCCGLRSTROARVVGGDADAGEPMQVSL 460

537 QCEDRSCTKKKPNPCDCKPDCRDSDEHCDGLG--PSSRIYGVGAVSSEGEPMQASL 594

461 HALGQGHICGASLSPNMLVSAHCYIDRGFRYSPTQWTAFLGLHDGQSRAPVOER 520

595 QVNGR-HICGGLIADRWVITAHCQED--SWASTVLTWTFVGLKWKQNSR-WPGEVSF 649

521 RLKRIISHPPFNDFPYDIALLEKPAEYSSMWPRICLPDASHVYPACKAIVTWGMH 580

650 KVSRLLAHPHEBDSHDYDVALQLDHPVRSAAVRVCLPARSHPEPDLHCMTWGMGA 709

581 TOYGGTALILQGEIRIVNQTCEMLPQOITPRMNCVFLSGVDSCGSDGSPLSV 640

710 LRESGPLSNALQKXVDVLIQDLCSEVRYQVTPRMLCAGRYGKGDACGSDGSPLYCK 769

641 EADGRIFQAGVSWGDCAQRNRPVYTRLLPLFRDWIKE 679

770 ALSGRWFLAGLVSWGLGCGRPYFGVYTRITGVISWIOQ 808

```
RESULT 16
O6UXD8 PRELIMINARY; PRT; 802 AA.
ID O6UXD8;
AC O6UXD8;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE PVAE354.
GN ORFNames=UNO354;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22887296; PubMed=12975309;
RA Clark H.F., Gurney A.L., Abaya E., Baker K., Baldwin D., Brush J.,
RA Chen J., Chow B., Chui C., Crowley C., Currell B., Denel B., Dowd P.,
RA Eaton D., Foster J., Grimaldi C., Gu Q., Hass P.B., Heldens S.,
RA Huang A., Kim H.S., Klimowski L., Jin Y., Johnson S., Lee J.,
RA Lewis L., Liao D., Mark M., Robbie E., Sanchez C., Schoenfeld J.,
RA Seshagiri S., Simmons L., Singh J., Smith V., Stinson J., Vagts A.,
RA Vandlen R., Watanabe C., Wieand D., Woods K., Xie M.H., Yansura D.,
RA Yi S., Yu G., Yuan J., Zhang M., Zhang Z., Goddard A., Wood W.I.,
RA Godowski P.;
RT "The secreted protein discovery initiative (SPDI), a large-scale
RT effort to identify novel human secreted and transmembrane proteins: a
RT bioinformatics assessment.";
RL Genome Res. 13:2265-2270(2003).
CC -1- SIMILARITY: Belongs to peptidase family S1.
DR EMBL; AY358398; AAQ88764.1; -.
DR InterPro; IPR000859; CUB.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR InterPro; IPR009003; Pept_Ser_Cys.
DR Pfam; PF00431; CUB; 1.
DR Pfam; PF00057; Ldl_recept_a; 2.
DR Pfam; PF00089; Trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00261; LDLRECEPTOR.
DR SMART; SM00192; LDLA; 3.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS01180; CUB; 1.
DR PROSITE; PS01209; LDLR_1; 1.
DR PROSITE; PS50068; LDLR_2; 3.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KM Hydrolyase; Protease; Serine protease.
SQ SEQUENCE 802 AA; 88845 MW; 8726C91B1E02E163 CRC64;

Query Match 26.5%; Score 1005; DB 2; Length 802;
Best Local Similarity 34.0%; Pred. No. 3.1e-62;
Matches 217; Conservative 98; Mismatches 230; Indels 94; Gaps 17;
```

```
282 PGQFTCRGCIKREKCDGMADCTDHSDELNCSCDAGHQTCKNFKCPKLFMVCDSDVND 341
450 RGEFLCSV-----NGLCVP---ACDGVKD 470
OY 342 CGNSDBQCGSCPAQIFRC-SNGKCLSKSQCGKDDCGDSDPACSPKVVYTCRHTY 400
DB 471 CPNGLBERNCVCA--PFQCKEDSTICSLPVCVCGQDCLNGSDEOCQ--GVPCTFTF 527
OY 401 RCNGLCLSKNCEQCKEDCSGDSKDCDGLRSTFQARVVGTDADGEMPMQVSL 460
DB 528 QCEDRCVCKPNQCGRPDRCRGSDEHDCGLQ--PSSRLVGAASSEGMPQASL 565
OY 461 HALGQGHICASLISPNMIVSAACVYIDRGFRYSPTQWTAFLGLHDSQRSAPGVQR 520
DB 586 QVRGR-HICGALILADRWVITAAHCFQED---SMASTVMTVFLGKVMQNSR-WPGEVSR 640
OY 521 RLKRIISHPPENDFTDYDIALLEKPAEYSSMVRPICLPDASHPPAKAIWYGMGH 580
DB 641 KVRRLIHPYHEBDSHDYDVALQLDHPVRSAAVPCVLPASHPEEPGLHGMWGMA 700
OY 581 TOYGGTALILQGEIRVINGTCENTLPPQITPRMVCVFLSGVDSQCGDSGFLSV 640
DB 701 LRGGPISNALQKVYDQLIPDLCSAARYQVTPRILCAGYRKCKDADCGDSGFLVCK 760
OY 641 EADGRIFQAGVSWDGCACQNRKPGVYTRLPFRDWIKE 679
DB 761 ALSGRMFLAGLVSWGLGCGRPNVFGVYTRITGVISMIQ 799

RESULT 17
AAQ88764 PRELIMINARY; PRT; 802 AA.
ID AAQ88764;
AC AAQ88764;
DT 02-MAR-2004 (TREMBlrel. 27, Created)
DT 02-MAR-2004 (TREMBlrel. 27, Last sequence update)
DT 02-MAR-2004 (TREMBlrel. 27, Last annotation update)
DE PVAE354.
GN UNO354.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primata; Catarrhini; Hominidae; Homo.
OC NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=12975309;
RA Clark H.F., Gurney A.L., Abaya E., Baker K., Baldwin D., Brush J.,
RA Chen J., Chow B., Chui C., Crowley C., Currell B., Denel B., Dowd P.,
RA Eaton D., Foster J., Grimaldi C., Gu Q., Hass P.B., Heldens S.,
RA Huang A., Kim H.S., Klimowski L., Jin Y., Johnson S., Lee J.,
RA Lewis L., Liao D., Mark M., Robbie E., Sanchez C., Schoenfeld J.,
RA Seshagiri S., Simmons L., Singh J., Smith V., Stinson J., Vagts A.,
RA Vandlen R., Watanabe C., Wieand D., Woods K., Xie M.H., Yansura D.,
RA Yi S., Yu G., Yuan J., Zhang M., Zhang Z., Goddard A., Wood W.I.,
RA Godowski P.;
RT "The Secreted Protein Discovery Initiative (SPDI), a Large-Scale
RT Effort to Identify Novel Human Secreted and Transmembrane Proteins: A
RT Bioinformatics Assessment.";
RL Genome Res. 13:2265-2270(2003).
DR EMBL; AY358398; AAQ88764.1; -.
SQ SEQUENCE 802 AA; 88845 MW; 8726C91B1E02E163 CRC64;

Query Match 26.5%; Score 1005; DB 2; Length 802;
Best Local Similarity 34.0%; Pred. No. 3.1e-62;
Matches 217; Conservative 98; Mismatches 230; Indels 94; Gaps 17;
```

```

Qy 170 GLRKAQSTFNSPYRGHYPPNIDCTWNIIEVPPNQHVKRFEFYLLPRRACGTCPKDY 229
Db 333 DNRLDSDQVLSIPYFSTYSPOTHCMSHLTVPSLDYGLALFADYALRRQKIDLPCTGQ 392
Qy 230 VEINGEKYCG-----ERSQFVVTNSNKITVRFHSDOSYTDGTFLAEYLSYDSDPC 281
Db 393 WTIGNRRLCGRLIQPYAERIPVAVTAG---ITINFTSQISLTGPGVAVHGLVNGSDPC 449
Qy 282 PGQFTCTGRCIRKELRCDCGMADCTDHSDELNCSCDAGHQTCKNKKCKPLFWVCDVND 341
Db 450 PGEFLCSV-----NGLCVP---ACDGVKD 470
Qy 342 CGDSDSDQSCSPAQTRFC-SNGKCLSKSQCNKDKDGDGSDASCPKANVVTCTKHTY 400
Db 471 CPNGLDERNCVCRA-TFOCKEDSTCISLPKVCDDQPDCLNGSDEQCE--GVPCGTFTF 527
Qy 401 RCLNGCLSKNPECDKEDSDGSDKDCDGLRSTFROARVVGTDADGEWPMQVSL 460
Db 528 QCEBRSQVKNPQCDGRPCDRSDSDEHCDGLOG--PSRITVGAVSSGSEWPMQASL 585
Qy 461 HALQGHICGASLISPNVLVSAHCYIDRGRFRYSDPQTWTAFLGLHDQSORSAFQVER 520
Db 586 QVRGR-HICGALLADRWVITPAHCFQED---SMASITVLTWFLGKVMQNSR-WPGEVSF 640
Qy 521 RLKRIISHPFNDTFEDYDIALLEKPAEYSSWVRPCLPDASHVFPAGKAIWWTGMGH 580
Db 641 KVSRLLPHPHEEDSHDYVALQLDHPVRSAAVRPCLPARSHFEPGLHCWITWGGA 700
Qy 581 TOYGGTGA-LIIOKGEIRVINOTTCENTLLPQOITPRPMCVGFLSGVDSQCGSGPLSSV 640
Db 701 LRBSGPISNALQKVQVLIPODLCEAVRYQVTPPMLCAGRKKGKDAQCGDSGSLVCK 760
Qy 641 EADGRIFQAGVSWGDCGAQRNKGVTYRLPLFRDWIK 679
Db 761 ALSGRMFLAGLVSMGLGCGRPYFGVYTRITGVISMIQ 799

RESULT 18
ID Q6IC2 PRELIMINARY; PRT; 824 AA.
AC Q6IC2;
DT 05-JUN-2004 (TREMBlrel. 27, Created)
DT 05-JUN-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUN-2004 (TREMBlrel. 27, Last annotation update)
DE D11170K4.2 protein.
GN Name=d11170K4.2;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Collins J.E., Wright C.L., Edwards C.A., Davis M.P., Grinham J.A.,
RA Cole C.G., Goward M.E., Aguado B., Malaya M., Mokrab Y., Huckle E.J.,
RA Beare D.M., Dunham I.,
RA Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
CC -1 SIMILARITY: Belongs to peptidase family S1.
EMBL; CRA56446; CAG30332.1; -.
DR InterPro: IPR000859; CUB.
DR InterPro: IPR001254; Peptidase_S1.
DR InterPro: IPR001314; Peptidase_S1.
DR InterPro: IPR009003; Pept_Ser_Cys.
DR Pfam: PF00431; CUB; 1.
DR Pfam: PF00057; Ldl_recept_a; 2.
DR Pfam: PF00089; Trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR PRINTS: PR00261; LDRRECEPTOR.
DR SMART; SM00192; Ldla; 3.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS01180; CUB; 1.
DR PROSITE; PS01209; LdLRA_1; 1.

```

```

DR PROSITE; PS0068; LDLRA_2; 3.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Protease; Serine protease.
SQ SEQUENCE 824 AA; 9133 MW; A74F186406041F7B CRC64;

Query Match 26.0%; Score 985; DB 2; Length 824;
Best Local Similarity 32.8%; Pred. No. 8.2e-61;
Matches 217; Conservative 98; Mismatches 230; Indels 116; Gaps 18;

Qy 67 PAH--ARCOMALRGDADSVLSLTFRSFDLASCDERGSDDLVTVYNTLSPMMEHALVOL--C 122
Db 229 PDHLASCLMHLQPKMLKRL--EWTLAEGR--LAMVAVAGLEKRLITLTVYGC 283
Qy 123 GTYPPSTNLT-----FHSSONVLLITLINTERRHPGEATFPQLFRMSCG 169
Db 284 SRQPVVSEVLASGAIMVWKKGLHSYYDPVLSV-----QPVVFQACEVNLTL 332
Qy 170 GLRKAQSTFNSPYRGHYPPNIDCTWNIIEVPPNQHVKRFEFYLLPRRACGTCPKDY 229
Db 333 DNRLDSDQVLSIPYFSTYSPOTHCMSHLTVPSLDYGLALFADYALRRQKIDLPCTGQ 392
Qy 230 VEINGEKYCG-----ERSQFVVTNSNKITVRFHSDOSYTDGTFLAEYLSYDSDPC 281
Db 393 WTIGNRRLCGRLIQPYAERIPVAVTAG---ITINFTSQISLTGPGVAVHGLVNGSDPC 449
Qy 282 PGQFTCTGRCIRKELRCDCGMADCTDHSDELNCSCDAGHQTCKNKKCKPLFWVCDVND 341
Db 450 PGEFLCSV-----NGLCVP---ACDGVKD 470
Qy 342 CGDSDSDQSCSPAQTRFC-SNGKCLSKSQCNKDKDGDGSDASCPKANVVTCTKHTY 400
Db 471 CPNGLDERNCVCRA-TFOCKEDSTCISLPKVCDDQPDCLNGSDEQCE--GVPCGTFTF 527
Qy 401 RCLNGCLSKNPECDKEDSDGSDKDCDGLRSTFROARVVGTDADGEWPMQVSL 460
Db 528 QCEBRSQVKNPQCDGRPCDRSDSDEHCDGLOG--PSRITVGAVSSGSEWPMQASL 585
Qy 528 QCEBRSQVKNPQCDGRPCDRSDSDEHCDGLOG--PSRITVGAVSSGSEWPMQASL 585
Qy 461 HALQGHICGASLISPNVLVSAHCYIDRGRFRYSDPQTWTAFLGLHDQSORSAFQVER 520
Db 586 QVRGR-HICGALLADRWVITPAHCFQED---SMASITVLTWFLGKVMQNSR-WPGEVSF 640
Qy 521 RLKRIISHPFNDTFEDYDIALLEKPAEYSSWVRPCLPDASHVFPAGKAIWWTGMGH 580
Db 641 KVSRLLPHPHEEDSHDYVALQLDHPVRSAAVRPCLPARSHFEPGLHCWITWGGA 700
Qy 581 TOYGGTGA-----LIIOKGEIRVINOTTCENTLLPQOITPRPMCVGFLSGVDSQCGSGPLSSV 618
Db 701 LRBSGPISNALQKVQVLIPODLCEAVRYQVTPPMLCAGRKKGKDAQCGDSGSLVCK 760
Qy 619 VGLSGVDSQCGSDGSGPLSSVEADGRIFOAGVSWGDCGAQRNKGVTYRLPLFRDWIK 678
Db 761 AGYRKKGKDAQCGSDGSGPLVCKALSGRMFLAGLVSMGLGCGRPYFGVYTRITGVISMIQ 820
Qy 679 E 679
Db 821 Q 821

RESULT 19
ID CAG30332 PRELIMINARY; PRT; 824 AA.
AC CAG30332;
DT 01-JUN-2004 (TREMBlrel. 27, Created)
DT 01-JUN-2004 (TREMBlrel. 27, Last sequence update)
DT 01-JUN-2004 (TREMBlrel. 27, Last annotation update)
DE D11170K4.2 protein.
GN D11170K4.2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
CX NCBI_TaxID=9606;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RA Collins J.E., Wright C.L., Edwards C.A., Davis M.P., Grinham J.A.,
RA Cole C.G., Goward M.E., Aguado B., Mallya M., Mokrab Y., Huckle E.J.,
RA Baere D.M., Dunham I.
RU Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL, CR456446; CAG30332.1; -.
SQ SEQUENCE 824 AA; 91333 MW; A74F186406041F7B CRC64;

Query Match 26.0%; Score 985; DB 2; Length 824;
Best local Similarity 37.8%; Pred. No. 8,2e-61;
Matches 217; Conservative 98; Mismatches 220; Indels 116; Gaps 18;

QY 67 PAH--ARQVALRGDADSVSLTFRSPDLASCDERGSGLVTVYNTLSMPMEHALVQD--C 122
DB 229 PDHLASCLMHLQPKDMLMLRL--EWTLAECRDR---LAMVDVAPLEKRLITSYGC 283
QY 123 GTYPPSTNLT-----FHSSONVLLTLINTERRHGFATPQLPRMSSCG 169
DB 284 SROEPVEVLASGAIMAVWKKGLHSYDPEVLSV-----QPVVFOACEVNLTL 332
QY 170 GRLKAQGTENSPYPPGHYPNIDCTWNIIEVNNQHVKNRKFYLLLEPRACGTCPKDY 229
DB 333 DNRLLDSQGLSTYFPSTYSPTQTCSMHLTVPSLDYGLALMFDYALRRQKTYDLPTQCG 392
QY 230 VEINGEKRYCG-----ERSQFVYTSNSMKITVRFHSDSYDTGFLAEYLSYSSDPG 281
DB 393 WTIGNRRLCGRLIHPYAEIRIPVAVATG---ITINFTSQISLTGPVAVHGLYNQSDPC 449
QY 282 PGQCTCTGRCIRKELACDGMADCTDHSDELNCSGADGHQPTCKNCKPLFWCDSVND 341
DB 450 PGEELCSV-----NGLCVP---ACDGVKD 470
QY 342 CGADSDSDQSCSPAQTRFC--SNGKCLSKSQCGKNGDGDSDBASCPKVNVVTCTGHTY 400
DB 471 CPNGLDERNCVCRA--TQCKEDSTCISLPKVCDDQPPCLNGBDEBQGF--GVPCGTF 527
QY 401 RCLNGCLSKNPECDKEDCDSDSKDCDGLRSTFRQAVVGTADGEMVPCVSL 460
DB 528 QCEDRSQVCKKPNFCDDRPDCRDSDBEHCCGLOG--PSSRIYGAIVSSBGEVPMQASL 585
QY 461 HALGQGHICGASLISPMVLSAACHYIDDKGRFRSDTQMTATPLGLDQSRASPGQER 520
DB 586 QVRER--HICGGLADRWVITTAACFOED--SNASTVLMVTFGKWKQNSR--WPGVSP 640
QY 521 RLKRIISHPFENDTCTFYDIALLELEKPAEYSSWVRPCLPDASHVPAGKAIWVTGWH 580
DB 641 KVSRLLLHPHREBDSHYDVALLDLHPVRSAAVRVCLPARSHFEPGLHCHITWGGA 700
QY 581 TQYGGTGA-----LIIQGEIRVYNQTTGENTLLPQGITPRMMC 618
DB 701 LREGALRADAVALEFYGWRNQGSETECCPISNALQKVNVQLPOLCSEVYRYQTTPMLC 760
QY 619 VGFPLSGVDSCQGGSGPLSVEADGRIFQAGVYSMDGCAQRKRPVYTRLPFRMWIK 678
DB 761 AGYRKGGKDACQGGSGPLVCALKASGRWFLAGLVSWGLGCGRPVYFGVYTRITGVISMIQ 820
QY 679 E 679
DB 821 Q 821

RESULT 20
ENTR_PIG STANDARD; PRT; 1034 AA.
AC P98074;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Enteropeptidase precursor (EC 3.4.21.9) (Enterokinase).
GN Name=PRSS7; Synonyms=ENTK;
OS Sus scrofa (Pig);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RU TISSUE=Duodenal mucosa;
RX MEDLINE=94327548; PubMed=8051081;
RA Matsushima M., Ichinose M., Yahagi N., Kakei N., Tsukada S., Miki K.,
RA Kurokawa K., Tashiro K., Shiohara K., Shinomiya K., Uneyama H.,
RA Inoue H., Takahashi T., Takahashi K.;
RT "Structural characterization of porcine enteropeptidase.";
RL J. Biol. Chem. 269:19976-19982(1994).
CC -I- FUNCTION: Responsible for initiating activation of pancreatic
CC proteolytic proenzymes (trypsin, chymotrypsin and carboxypeptidase
CC A). It catalyzes the conversion of trypsinogen to trypsin which in
CC turn activates other proenzymes including chymotrypsinogen,
CC procarboxypeptidases, and proelastases.
CC -I- CATALYTIC ACTIVITY: Selective cleavage of 6-Lys-|-Ile-7 bond in
CC trypsinogen.
CC -I- SUBUNIT: Heterotrimer of a catalytic (light) chain, a multidomain
CC (heavy) chain, and a mini chain.
CC -I- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
CC -I- The chains are derived from a single precursor that is
CC cleaved by a trypsin-like protease.
CC -I- PFM: The mini chain may be cleaved by elastase.
CC -I- SIMILARITY: Belongs to peptidase family S1.
CC -I- SIMILARITY: Contains 2 CUB domains.
CC -I- SIMILARITY: Contains 2 LDL-receptor class A domains.
CC -I- SIMILARITY: Contains 1 MAM domain.
CC -I- SIMILARITY: Contains 1 SEA domain.
CC -I- SIMILARITY: Contains 1 SRCR domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, D30799; BAA06459.1; -.
DR HSSP, P98072; 1EBX.
DR MR08B, S01.156; -.
DR InterPro: IPR000859; CUB.
DR InterPro: IPR002172; LDL_receptor_A.
DR InterPro: IPR000998; MAM.
DR InterPro: IPR001254; Peptidase_S1.
DR InterPro: IPR001314; Peptidase_S1A.
DR InterPro: IPR011163; Pept_S1A_enterop.
DR InterPro: IPR009003; Pept_Ser_Cys.
DR InterPro: IPR000082; SEA.
DR InterPro: IPR001190; Strc_receptor.
DR Pfam, PF00431; CUB; 2.
DR Pfam, PF00057; Ldl_recept_a; 2.
DR Pfam, PF00629; MAM; 1.
DR Pfam, PF01390; SEA; 1.
DR Pfam, PF00530; SRCR; 1.
DR Pfam, PF00089; Trypsin; 1.
DR PIRSF, PIRSF001138; Enteropeptidase; 1.
DR PRINTS, PR00722; CHYMOTRYPSIN.
DR PRINTS, PR00261; LDLRECEPTOR.
DR PRINTS, PR00020; MAMDOMAIN.
DR SMART, SM00042; CUB; 2.
DR SMART, SM00192; LDLa; 2.
DR SMART, SM00137; MAM; 1.
DR SMART, SM00200; SEA; 1.
DR SMART, SM00202; SR; 1.
DR SMART, SM00020; Tryp_SPC; 1.
DR PROSITE, PS01180; CUB; 2.
DR PROSITE, PS01209; LDLa; 1; 2.
DR PROSITE, PS50068; LDLa_2; 2.
DR PROSITE, PS00740; MAM_1; 1.
DR PROSITE, PS50060; MAM_2; 1.
DR PROSITE, PS50024; SEA; 1.

```

```

DR PROSITE; PS00420; SRCR_1; FALSE_NEG.
DR PROSITE; PS50287; SRCR_2; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Direct protein sequencing; Glycoprotein; Hydrolyase; Lipoprotein;
KW Myristate; Repeat; Serine protease; Signal-anchor; Transmembrane;
KW Zymogen.
FT CHAIN 52 117 Enteropeptidase non-catalytic mini chain.
FT CHAIN 118 799 Enteropeptidase non-catalytic heavy
FT CHAIN chain.
FT CHAIN 800 1034 Enteropeptidase catalytic light chain.
FT DOMAIN 1 18 Cytoplasmic (Potential).
FT TRANSMEM 19 47 Signal-anchor for type II membrane
FT protein (Potential).
FT extracellular (Potential).
FT SEA.
FT LDI-receptor class A 1.
FT CUB 1.
FT MAM.
FT CUB 2.
FT LDI-receptor class A 2.
FT SRCR.
FT Serine protease.
FT Charge relay system (By similarity).
FT ACT_SITE 840 840 Charge relay system (By similarity).
FT ACT_SITE 891 891 Charge relay system (By similarity).
FT ACT_SITE 986 986 Charge relay system (Potential).
FT LDI 2 2 N-myristoyl glycine (Potential).
FT DISULFID 199 212 By similarity.
FT DISULFID 206 225 By similarity.
FT DISULFID 219 236 By similarity.
FT DISULFID 658 670 By similarity.
FT DISULFID 665 683 By similarity.
FT DISULFID 677 692 By similarity.
FT DISULFID 787 911 Interchain (By similarity).
FT DISULFID 825 841 By similarity.
FT DISULFID 925 992 By similarity.
FT DISULFID 956 971 By similarity.
FT DISULFID 982 1010 By similarity.
FT CARBOHYD 116 116 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 147 147 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 170 170 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 194 194 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 283 283 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 343 343 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 350 350 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 403 403 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 455 455 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 485 485 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 518 518 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 549 549 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 645 645 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 697 697 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 701 701 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 721 721 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 740 740 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 761 761 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 804 804 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 863 863 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 902 902 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 964 964 N-linked (GlcNAc... ) (Potential).
SQ SEQUENCE 1034 AA; 114776 MW; 0386C64CF6CC368 CRC64;

Query Match 19.0%; Score 718.5; DB 1; Length 1034;
Best Local Similarity 32.4%; Pred. No. 6,1e-42;
Matches 181; Conservative 87; Mismatches 189; Indels 101; Gaps 22;

```

```

Db 591 -----IADVBEIRDEEDSLILAVYTGPGVEDVSTTNMTVLFTINDALTKGFK 643
QY 270 AEYLS---YDSDDPC-PGQFTCRGRCIRKELRCGDWADCTDHSDELNCSGDACHQFTCK 325
Db 644 ANFTGYHLGIPECKEDNFCQENGCVCVLLVNLCDGFSCKDSDENACV----- 693
QY 326 NKPEKPLFWCVDSYVDCGDSDEGSCSPAQTRFCS---NGKLSKSQCCNGKQDC---G 379
Db 694 -RF-----LNGRANNSG-----LVQFRIGSIWHTACAEWTTQTSDVQCGLG 735
QY 380 DGSDEASCP-----KVVVYTCYKHTYRCLNGLCSKGNBECDKED-----CSD 423
Db 736 LGTNSSMPEFSSGSGGPFVLTNP-----NGSLITLASPOC--FEDSLILQCN 784
QY 424 GSDKDDDCGLRSTFRA--RVVGCTDADGEMPMQVSLHALGQHI CGASLSPMLVS 481
Db 785 KS-----CGKQVQSVSPRIVGNSRSGAMPMVVALYYNQ-LICGASLSRDLVS 837
QY 482 AAHCYIDRGRFYSDPTQWTAFLGLHDQSOSAAGVQERRIKRTIISHPFNDPTFDYDIA 541
Db 838 AAHCYVG---RNLPEKMKAILGLHMTSNLTSPOIYTRILDEIVINPHYNRKQSDIA 893
QY 542 LLELEKPAEYSSWVRPILCPDASHVPPAGAIWVGHTQYGGTALLIQGEIRVINO 601
Db 894 MHLFEFVNTYDYOPICLPEENQVFPGRIGSIAGKVIYQSSPADIIQEDADVPLLSN 953
QY 602 TTGCENLDPQ-QIFRMMCVGFLSGVDS CGDSSGSPISVADGRIRPQGVSMGDCAQ 660
Db 954 EKCQQQPEVNTIENMNCAGIEBGIDSCGDSGGLMCLN-NRMWLAGVTSFGYCAL 1012
QY 661 RNKPGVYTRLPFRDWIK 678
Db 1013 PNRGVYARVPKFTWIO 1030

RESULT 21
Q72410 PRELIMINARY; PRT; 855 AA.
ID 072410
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Polyease-1B protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22784742; PubMed=12886014;
RA Cal S., Lopez-Otin C.;
RT "Polyeasase-1, a human polypeptide with the ability to generate
RT independent serine protease domains from a single translation
RT product."
RL Proc. Natl. Acad. Sci. U.S.A. 100:9185-9190 (2003).
DR EMBL; AJ488947; CAD35759.1; -.
DR MEROPS; S01.357; -.
DR MEROPS; S01.358; -.
DR MEROPS; S01.969; -.
DR GO; GO:0004295; F:Trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR001172; LDL_receptor_A.
DR InterPro; IPR001254; Peptidase_S1.
DR Pfam; PF00087; Ldl_recept_a; 1.
DR SMART; SM00192; Ldla; 1.
DR SMART; SM00020; Tryp_SPC; 2.
DR PROSITE; PS50240; TRYPSIN_DOM; 2.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_2.
DR PROSITE; PS00135; TRYPSIN_SER; 2.
KW Hydrolyase; Protease; Serine protease.

```

SEQ	SEQUENCE	855 AA;	91352 MW;	8AR2759D9740CF3F CRC64;
Query Match	18.9%;	Score 717.5;	DB 2;	length 855;
Best Local Similarity	45.7%;	Pred. No. 5.8e-42;		
Matches 132;	Conservative 53;	Mismatches 95;	Indels 9;	Gaps 5;
Qy	CTKHTRYRLNGLCLSGKGNPECGKXKDCSDGSGEXDCDGLASFPTQA-RVYCGTDADEGE	453		
Db	154 CPGNSFSGNSOCVTKANPBCDDQEDSCSDSEAHCEGLQAWMAGHIVGEMASPEE	213		
Qy	454 WPMOVSIHALQGHICGASLISPMNLVSAAHCYIDRGFRYSDPTQMTAFLGLHDQGRS	513		
Db	214 FPMQASLRE-NKEHFCGAIIINARLTVSAHCFNE-----FDDPTKWAIVAYGATLISGE	267		
Qy	514 APQVQERLKIIISHPPNDFEDYDIALLELEKPAEYSSMWPRICLPDASHVFPAGKAI	573		
Db	268 ASTVA-AQVAVIVKPIPLYNADPADDEVAVLELTSPFCGRHIQVPLCPAPATHIPEPSKCC	326		
Qy	574 WVTGWHGHTQYGG-TCALTLQKEIRIVINOTTCEMLLPQIIPRMNCVGPISGVNSCGD	632		
Db	327 LISGEGYUKEDPLVYKEVLAQKAYVELLDQALCASLYGHSILTRMVCAGLIDKVDSCGD	386		
Qy	633 SCGPLLSVEADGRIFQAGVNSWGDCQAQRNKRGVYTRPLPLFDWIKENT 681			
Db	387 SCGPLVCEBPSGRFTLACIVSWGICACAAARRRGVATYTRLRDLWILEAT 435			

Query Match	18.9%	Score 717.5	DB 2	Length 1059
Best local Similarity	45.7%	Pred. No. 7.3e-42		
Matches 132	Conservative 53	Mismatches 95	Indels 9	Gaps 5

```

395 CTGATTTCATGACATGSKNPNEDGKDCGDSGDEKCCDGGASFTTQA-RVYGGTDADGCE 453
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

Db 154 CPENSPSCNSQCVTKNPECDQDSCDSGDEAHCEGLOPAWRMAGRIYVGMEASPG 213

Qy 454 MNPQVSLHALGGGHICGASLISPMNLVSAACIYIDBGRFSPTQYTAFLGHDGORS 513

Db 214 FPMQSLBRE-NNEHFCGAIIINARVLASAHCENE-----FQDPTKVAVAATYLSGE 267

Qy 514 APVQSERRLKRIISHPFENDFPDVILBLEKPAEYSWVRICLPDASHVFPAGKAI 573

Db 268 ASIVYR-AQVVOIVGHPILNADPDVAIVLEISLPFGRIHICVCLPAATHIFPSPKCC 326

Qy 574 WYTGWHTGYGG-TGALLILQKGEIRVINQTTCEMLPQGITPRMVCVGLSGGVDSQGD 632

Db 327 LISGMGYLKEDELVPVEVLQKATVEILDQALCASLYGSLIDRWVCAGYLDGKVDSCGD 386

Qy 633 SGGSPSSVSEADGRIFQAGVSWSGSCAORNKRGVYTLPLFRMDIKENT 681

Db 387 SGGPLVCEBSPGRFPLAGVSWGLGCAARRPGVYARTLRDLIDILEAT 435

```

RESULT 23
ENTR_BOVIN          STANDARD;          PRT; 1035 AA.
ID   ENTR_BOVIN
AC   P98072;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-OCT-2004 (Rel. 45, Last annotation update)
DE   Enteropeptidase precursor (EC 3.4.21.9) (Enterokinase) .
GN   Name=PRSS7; Synonyms=ENTK;
OS   Bos taurus (Bovine) .
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC   Bovinae; Bos.
CX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=Ductenum;
RC   MEDLINE=94329561; PubMed=8052624;
RA   Kitamoto Y., Yuan X., Wu Q., McCourt D.W., Sadler J.E.;
RT   "Enterokinase, the initiator of intestinal digestion, is a mosaic
RT   proc. composed of a distinctive assortment of domains." ;
RL   Proc. Natl. Acad. Sci. U.S.A. 91:7588-7592(1994) .
RN   [2]
RP   SEQUENCE OF 801-1035 FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=94043122; PubMed=8226655;
RA   Lavallie E.R., Rehemullia A., Racie L.A., Dibiaseo E.A., Ferez C.,
RA   Grant K.L., Light A., McCoy J.M.;
RT   "Cloning and functional expression of a cDNA encoding the catalytic
RT   subunit of bovine enterokinase." ;
RL   J. Biol. Chem. 268:23311-23317(1993) .
RN   [3]
RP   SEQUENCE OF 801-827.
RC   TISSUE=Intestine;
RX   MEDLINE=92189715; PubMed=1799406;
RA   Light A., Janaka H.;
RT   "The amino-terminal sequence of the catalytic subunit of bovine
RT   enterokinase." ;
RL   J. Protein Chem. 10:475-480(1991) .
CC   -1- FUNCTION: Responsible for initiating activation of pancreatic
CC   proenzyme proenzymes (trypsin, chymotrypsin and carboxypeptidase
CC   A). It catalyzes the conversion of trypsinogen to trypsin which in
CC   turn activates other proenzymes including chymotrypsinogen,
CC   procarboxypeptidases, and proelastases.
CC   -1- CATALYTIC ACTIVITY: Selective cleavage of 6-hys-|-ile-7 bond in
CC   trypsinogen.
CC   -1- SUBUNIT: Heterodimer of a catalytic (light) chain and a
CC   multidomain (heavy) chain linked by a disulfide bond.
CC   -1- SUBCELLULAR LOCATION: Type II membrane protein (Probable) .
CC   -1- ALTERNATIVE PRODUCTS:
CC   Event=Alternative splicing; Named isoforms=2;
CC   Name=Long;
CC   IsoId=P98072-1; Sequence=Displayed;
CC   Name=Short;
CC   IsoId=P98072-2; Sequence=VSP_005386;
CC

```

CC -1- TISSUE SPECIFICITY: Intestinal brush border.
CC -1- PM: The chains are derived from a single precursor that is
CC cleaved by a trypsin-like protease.
CC -1- SIMILARITY: Belongs to peptidase family S1.
CC -1- SIMILARITY: Contains 2 CUB domains.
CC -1- SIMILARITY: Contains 2 LDL-receptor class A domains.
CC -1- SIMILARITY: Contains 1 SPA domain.
CC -1- SIMILARITY: Contains 1 SRCR domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for commercial
CC entities requires a license agreement (See [http://www.ebi.ac.uk/announcements/](http://www.ebi.ac.uk/announcements)
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; U09859; AAB40026.1; -;
DR EMBL; L19663; AAB40035.1; -;
DR PIR; A43090; A43090.
DR MEROPS; S01.156; -;
DR InterPro; IPR000859; CUB.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR000998; MAM.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR InterPro; IPR011163; Pept_S1A_enterop.
DR InterPro; IPR009003; Pept_Ser_Cys.
DR InterPro; IPR000082; SRA.
DR InterPro; IPR001190; Srcr_receptor.
DR Pfam; PF00431; CUB; 2.
DR Pfam; PF00057; Ldl_recept_a; 2.
DR Pfam; PF00629; MAM; 1.
DR Pfam; PF01390; SEA; 1.
DR Pfam; PF00510; SRCR; 1.
DR Pfam; PF00083; Trypsin; 1.
DR PIRSF; PIRSF001138; Enteropeptidase; 1.
DR PRINTS; PRO0722; CHYMOTRYPSIN.
DR PRINTS; PRO0261; LDLRECEPTOR.
DR PROSITE; PS01180; CUB; 2.
DR PROSITE; PS01209; LDLRA_1; 2.
DR PROSITE; PS50068; LDLRA_2; 2.
DR PROSITE; PS00740; MAM_1; 1.
DR PROSITE; PS50060; MAM_2; 1.
DR PROSITE; PS50024; SEA; 1.
DR PROSITE; PS00420; SRCR_2; 1.
DR PROSITE; PS50287; SRCR_1; FALSE_NEG.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW 3D-structure; Alternative splicing; Direct protein sequencing;
KW Glycoprotein; Hydrolase; Lipoprotein; Myristate; Repeat;
KW Serine protease; Signal-anchor; Transmembrane; Zymogen.
FT CHAIN 1 800 Enteropeptidase non-catalytic chain.
FT CHAIN 801 1035 Enteropeptidase catalytic chain.
FT DOMAIN 1 18 Cytoplasmic (Potential).
FT TRANSMEM 19 47 Signal-anchor for type II membrane
FT protein (Potential).
FT EXTRACELL 1 Extracellular (Potential).
FT DOMAIN 48 1035 SEA.
FT DOMAIN 54 169 LDL-receptor class A 1.
FT DOMAIN 197 238 CUB 1.
FT DOMAIN 240 350 MAM.
FT DOMAIN 358 520 CUB 2.
FT DOMAIN 540 650 LDL-receptor class A 2.
FT DOMAIN 657 695 SRCR.
FT DOMAIN 694 787 Serine protease.
FT DOMAIN 801 1035 Charge relay system (By similarity).
FT ACT_SITE 841 841 Charge relay system (By similarity).
FT ACT_SITE 892 892 Charge relay system (By similarity).
FT ACT_SITE 987 987 N-myristoyl glycine (Potential).
FT LIPID 2 2

FT DISULFID 199 212 By similarity.
FT DISULFID 206 225 By similarity.
FT DISULFID 219 236 By similarity.
FT DISULFID 659 671 By similarity.
FT DISULFID 666 684 By similarity.
FT DISULFID 678 693 By similarity.
FT DISULFID 788 812 Interchain (By similarity).
FT DISULFID 826 842 By similarity.
FT DISULFID 926 993 By similarity.
FT DISULFID 957 972 By similarity.
FT DISULFID 983 1011 By similarity.
FT CARBOHYD 116 147 N-linked (GlcNAc...)
FT CARBOHYD 147 170 N-linked (GlcNAc...)
FT CARBOHYD 170 170 N-linked (GlcNAc...)
FT CARBOHYD 194 194 N-linked (GlcNAc...)
FT CARBOHYD 233 233 N-linked (GlcNAc...)
FT CARBOHYD 263 263 N-linked (GlcNAc...)
FT CARBOHYD 264 264 N-linked (GlcNAc...)
FT CARBOHYD 404 404 N-linked (GlcNAc...)
FT CARBOHYD 456 456 N-linked (GlcNAc...)
FT CARBOHYD 486 486 N-linked (GlcNAc...)
FT CARBOHYD 519 519 N-linked (GlcNAc...)
FT CARBOHYD 550 550 N-linked (GlcNAc...)
FT CARBOHYD 646 646 N-linked (GlcNAc...)
FT CARBOHYD 698 698 N-linked (GlcNAc...)
FT CARBOHYD 722 722 N-linked (GlcNAc...)
FT CARBOHYD 741 741 N-linked (GlcNAc...)
FT CARBOHYD 762 762 N-linked (GlcNAc...)
FT CARBOHYD 864 864 N-linked (GlcNAc...)
FT CARBOHYD 903 903 N-linked (GlcNAc...)
FT CARBOHYD 965 965 N-linked (GlcNAc...)
FT VARSPIC 166 192 Missing (in isoform Short).
FT CONFLICT 808 808 /PRTD_VSP_005386.
FT STRAND 802 802 R -> Y (in Ref. 3).
FT TURN 805 806
FT TURN 809 810
FT TURN 813 814
FT STRAND 815 820
FT TURN 821 822
FT STRAND 823 830
FT STRAND 835 838
FT HELIX 840 843
FT TURN 844 845
FT HELIX 850 852
FT STRAND 853 857
FT TURN 861 861
FT STRAND 862 863
FT TURN 868 869
FT STRAND 871 880
FT TURN 882 883
FT STRAND 885 885
FT TURN 886 889
FT STRAND 890 890
FT STRAND 894 898
FT STRAND 912 912
FT TURN 916 917
FT TURN 922 923
FT STRAND 925 930
FT STRAND 933 933
FT TURN 936 937
FT STRAND 940 940
FT STRAND 943 943
FT STRAND 945 951
FT HELIX 954 960
FT TURN 962 963
FT TURN 968 969
FT STRAND 970 973
FT TURN 976 977
FT STRAND 981 981
FT STRAND 984 985
FT TURN 987 988
FT STRAND 990 995


```

FT DOMAIN 579 614 LDL-receptor class A 5.
FT DOMAIN 615 653 LDL-receptor class A 6.
FT DOMAIN 654 690 LDL-receptor class A 7.
FT DOMAIN 690 786 SRCR.
FT ACT_SITE 802 1042 Serine protease.
FT ACT_SITE 843 892 Charge relay system (By similarity).
FT ACT_SITE 892 892 Charge relay system (By similarity).
FT ACT_SITE 985 985 Charge relay system.
FT DISULFID 790 912 By similarity.
FT DISULFID 828 844 By similarity.
FT DISULFID 955 970 By similarity.
FT DISULFID 981 1010 By similarity.
FT CARBOHYD 80 80 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 104 104 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 135 135 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 141 141 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 231 231 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 245 245 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 251 251 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 305 305 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 320 320 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 376 376 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 413 413 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 446 446 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 451 451 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 469 469 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 567 567 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 651 651 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 697 697 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 761 761 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 1022 1022 N-linked (GlcNAc...) (Potential).
FT MUTAGEN 985 985 S->A: Loss of activity.
FT CONFLICT 854 854 W->R (in Ref. 2).
FT CONFLICT 876 876 K->R (in Ref. 2).
SQ SEQUENCE 1042 AA; 116564 MW; 7705398EBB607AD2 CRC64;

```

Query March Best Local Similarity 28.4%; Score 694; DB 1; Length 1042; Matches 214; Conservative 95; Mismatches 255; Indels 180; Gaps 34;

```

QY 30 DSKTVQRTDNDSGSGFLHARGVELMFTPGFPDPSYPYAHARCOMALRGDAD----- 81
DB 365 DHDCVDSDEVNCS--CHSGGLVECR-----NGCIPSTGQCD--GDBDCKDGDDEE 412
QY 82 --SYLSLTFNSFD-----LASCDEKSGDLVTVYNTL---SPMEPALVQLGCTYPPS 128
DB 413 NCSYIOTSQCGDRDCLYNPCLDSCG--GSSLCDPNNSLNCSOCEPITL-ELCMNLP-- 467
QY 129 YNLTFHSSQNVLLTLTLNTERRRHGFATPFQULPRMSSCGRLRK-AQGFNSPYVPGH 187
DB 468 YNST-----STPNY-----FGHRTQKESISWESSLPFAL 497
QY 188 YRPN-----IDCTWNIEVNNHVKVRFKFFLLBPRRACGTCPKDYVE-----INGEK 236
DB 498 VQTNCKYKLMFSCST--ILVE-----KCDVNTGERIRIPCRALCHSKKRCGSYVIGVLQ 550
QY 237 YCGER--SQPVVTSNSKITVRFHSDOSYTDGTGLAETLSYDSDPC-PGQFCTGTCGT 293
DB 551 WPEIDTDSQFPPEENSNDQTCI-----WPEDEV-----EECSFHFKRSQGCY 593
QY 294 RKELRCGGMACCTHSDSLNCSDAHQFTC-KKKFKCKFLFWVCDVVDGDNDEQGS 352
DB 594 LASRRCDGQACDDSDSENCCKERRDLMECPSNQCKLTKTIVICDGFPCDPYDEKNGS 653
QY 353 -CPAQTFRCNSGKCLSKSQCCNGKDDCCGSDDEASCPKVVN----- 392
DB 654 FCGDDELECAHACVSRDLWCDGEADCDSDSDMDCVTLSTINVASSSFLWVHRAATEHHV 713
QY 393 -----VTC-----TKRT-YRCLNGCLSK--GNP 413
DB 714 CADGMOBITLSQIAKQKMGIGEPSPVTKLIOQEKPERWLTLSHNSWESLGTTLAHLVNGQ 773
QY 414 ECDKEDKDSGSDSDKDCDCCGARSFTR-QARVVGCTDADDEGMPVQVSLHALGGCHIGAS 472

```

```

DB 774 SCESRSKISLTCMQ--DCGRPARARMKRIILGRTSRPGKWPQCISIQSPSGHICGV 831
QY 473 LISPNWLVSAACHYCIDDRGFRYSDPTQMTAFGLIHDSQSAPVQERRLKRIISHPFN 532
DB 832 LIARKWLVIAHCP-EER---ENAAVWKVVLGINNDHPSV-FMQTRPVKTLILHPRYS 885
QY 533 DFTFDYIALLELEKPAEVSMPVPCLPASHVFPAGKAIWYMGHTQYGGTALILQ 592
DB 886 RAVVDYISIVLESIEDISFTGYVAPVCLPNEQMLEPDYTCYITGMCH--MGNKMPKIQ 943
QY 593 KGEIRVNIQTCENILPQO--ITPRMCGFLSGGVDSGCGSGGLSSVADGRIFQAGV 651
DB 944 EGEVRILISLEHCQSPFMKITTIRMICAGYESSGTVDSGMDSGGSLVCEKRGKWLFG 1003
QY 652 VSWGDGAQR-NKRGVTRLPPLPRDWIK 678
DB 1004 TSWGVCFSKVLGPGVSNVSVEVEMWK 1031

```

```

RESULT 25
ID Q80YN4 PRELIMINARY; PRT; 1111 AA.
AC Q80YN4;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Corin.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley;
RA Langenickel T., Pagel I., Buttgeriet J., Jenner K., Lindner M.,
RL Willenbrock R., Dietz R., Bader M.;
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
CC 1- SIMILARITY: Belongs to peptidase family S1.
DB EMBL; AY251285; A086772.1; -.
DB HSBP; Q07954; ICR8.
DB GO; GO:0016020; C:membrane; IEA.
DB GO; GO:0044263; F:chymotrypsin activity; IEA.
DB GO; GO:0005044; F:scavenger receptor activity; IEA.
DB GO; GO:0004295; F:trypsin activity; IEA.
DB GO; GO:0007275; P:development; IEA.
DB GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DB InterPro; IPR000024; Pz domain.
DB InterPro; IPR002172; LDL_receptor_A.
DB InterPro; IPR001254; Peptidase_S1.
DB InterPro; IPR001314; Peptidase_S1A.
DB InterPro; IPR009003; Pept_Ser_Cys.
DB Pfam; PF01392; Fz; 2.
DB Pfam; PF00057; Ldl_recept_a; 6.
DB PRINTS; PR00722; CHYMOTRYPIN.
DB PRINTS; PR00261; LDLRECEPTOR.
DB SMART; SM00063; FRI; 2.
DB SMART; SM00192; IDLA; 7.
DB SMART; SM00202; SR; 1.
DB SMART; SM00020; TRYD_SRC; 1.
DB PROSITE; PS50038; Fz; 2.
DB PROSITE; PS01209; LDLRA_1; 6.
DB PROSITE; PS50068; LDLRA_2; 7.
DB PROSITE; PS50287; SRCR_2; 1.
DB PROSITE; PS50240; TRYPSIN_DOM; 1.
DB PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydroxylase; Protease; Serine protease.
SQ SEQUENCE 1111 AA; 122635 MW; 3BA2706CBB81157F CRC64;
Query March Best Local Similarity 18.2%; Score 690.5; DB 2; Length 1111;

```

Matches 210; Conservative 79; Mismatches 252; Indels 199; Gaps 32;

QY 33 TQVOTQNGSGFGLHARGVELMRTTGFDPSPYPARACQMALRGADVSLTFPSFD 92
 DB 462 TFCQDGEDCKDGDSENCSS-----DRPRP-----CPGDRGCLDS----- 497
 QY 93 LASGDE--RGSADVTVNTL---SPMEPALVQLCGTPSPYNTLTFH----- 134
 DB 498 --SVESECASSSLCDSSSLNSCHCEPITL-ELQNLPL--YNLTHPNVIGHRTQKEAS 552
 QY 135 -SSQNVLLTLLINTERRHPEFEATFPQLPRMSSCGRLRAQGTSPYPYHP-NT 192
 DB 553 ISMESALFPALVQNTCYKLMFFACTLLVPCKDVTNQ-----RVPPCRL 597
 QY 193 DCTWNIEVPPNNOHKVAFKFPYLLPERRAGTC-----PKDYVEINRGKVGGEESQF 244
 DB 598 LC-----EHSKER-----CESVIGVGLQMPED-----TDSQF 626
 QY 245 VVTSNKKITVRFHSDQSYTDGFLAAYLSYSSDPC-PCQFTGTCRIKELRCQDMA 303
 DB 627 PEGSSDNQTCL-----LPREDVEECSPSHFKRSRGCVLSGRCDGQA 669
 QY 304 DCTDHSDELNCSADAGHQFTCK-NKFCPLFWCDVNDGDNDEQCS-CPAQTRCS 361
 DB 670 DCDSDSDEBNCCKERDLMECPILKQCLKTLICDGPDCSDMDEKNCSPQDDELECA 729
 QY 362 NGKLSLSQCGKDDCGDSDSEASCRAVNYTCTKH-----TYRCING- 405
 DB 730 NHECVPRDLWCDGWTDCSDSDSEWGC-----VTLKNGNSSSPLVHRSARDHHVCADGW 784
 QY 406 -----LC--LSKGNPE---CDGKE-----DCSDGS 425
 DB 785 QETLSQACQMGEGPSVTELVQGGQGMRLHSSWMENLNGSTLQELVHRSCPSGS 844
 QY 426 DEK---DCDGLRSFTR-QARVVGCTDADGEMPMQVSLHALQGHICGASLISPMVLV 480
 DB 845 EISLCTKQCKGRPARMRKRIIGGRTSRGPRMWCOSLSEPSGHIICGVLLIAKKWVL 904
 QY 481 SAHCYCTDDGFRSDPTQWTAFLGLHDQGRSAPGVQERTKRIISHPPNDTPFYDI 540
 DB 905 TVAAHCF-EGR---EDADVNVVPGINNLDHPSC-FMQTRVKTLLHPKISRAVVIYDI 958
 QY 541 ALLELEPAEYSSWVRPCLPDASHVPAGKAVITWGHTQYGTGALLIQQKEIRVIN 600
 DB 959 SVVSLSDINDINETSIVRVCCLPSPREFLEPDYCYITWGH--MKNMKNPKLQBEBAVILIP 1016
 QY 601 QTTCENILLPQO-ITPRMNCVGLSGVDSQCGSGGGLSVYADGRIFQAGVSWSDGCA 659
 DB 1017 LEQCOSYFDMKTIITNMICAGYESGTVDSQWSDGGLVCEBPPGQWTLFGLTWGSVCF 1076
 QY 660 QR-NKPGVYTRLPFRDMIK 678
 DB 1077 SKVLGPGVYSNVSYFVDMIE 1096

RESULT 26
 O8CANG9 PRELIMINARY; PRT; 777 AA.
 AC O8CANG9;
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
 DE Mus musculus 16 days neonate thymus cDNA, RIKEN full-length enriched library, clone:Al30097D21 product:protease, serine, 7 (enterokinase), full insert sequence. (fragment).
 DE Name=Prs87;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP STRAIN=C57BL/6J; TISSUE=Thymus;
 RC

RX MEDLINE=99279253; PubMed=10349636;
 RA Carninci P., Hayashizaki Y.;
 RT "High-efficiency full-length cDNA cloning."; Meth. Enzymol. 303:19-44(1999).
 RL [2]
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Thymus;
 RX MEDLINE=21085660; PubMed=11217851;
 RA RIKEN FANTOM Consortium;
 RT "Functional annotation of a full-length mouse cDNA collection."; Nature 409:685-690(2001).
 RL [3]
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Thymus;
 RA The FANTOM Consortium.
 RT "Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs."; Nature 420:563-573(2002).
 RL [4]
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Thymus;
 RX MEDLINE=20499374; PubMed=11042159;
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M., Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes."; Genome Res. 10:1617-1630(2000).
 RL [5]
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Thymus;
 RX MEDLINE=20530913; PubMed=11076861;
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P., Kono H., Akiyama U., Nishi K., Kitsuai T., Tashiro H., Itoh M., Saito R., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kasaiwaagi K., Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Watanishi M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuyama S., Kawai U., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multiplexed capillary sequencer."; Genome Res. 10:1757-1771(2000).
 RL [6]
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Thymus;
 RA Adachi U., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P., Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W., Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T., Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T., Katoh H., Kawai U., Kojima Y., Kondo S., Kono H., Kouda M., Koya S., Kuribara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M., Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y., Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H., Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M., Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T., Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
 RT Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 RL [7]
 RN [7]
 RP SEQUENCE FROM N.A.
 CC -1- SIMILARITY: Belongs to peptidase family S1.
 CC -1- SIMILARITY: Contains 1 MAM domain.
 DR EMBL; AK038356; BAC29973.1; -.
 DR MGD; MGI:1197523; Prs87.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0004263; F:chymotrypsin activity; IEA.
 DR GO; GO:0016301; F:kinase activity; IEA.
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0005044; F:scavenger receptor activity; IEA.
 DR GO; GO:0004295; F:trypsin activity; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000859; CUB.
 DR InterPro; IPR002172; IDL_receptor_A.
 DR InterPro; IPR000998; MAM.
 DR InterPro; IPR001254; Peptidase_S1.
 DR InterPro; IPR001314; Peptidase_S1A.

DR InterPro: IPR009003; Pept. Ser. Cys.
 DR InterPro: IPR001190; Ser. Receptor.
 DR Pfam: PF00431; CUB; 2.
 DR Pfam: PF00057; Ldl_recept_a; 1.
 DR Pfam: PF00629; MAM; 1.
 DR Pfam: PF00530; SRCR; 1.
 DR Pfam: PF00089; Trypsin; 1.
 DR PRINTS: PR00722; CHYMOTRYPSIN.
 DR PRINTS: PR00020; MAMOMAIN.
 DR SMART: SM00042; CUB; 1.
 DR SMART: SM00192; LDLa; 1.
 DR SMART: SM00137; MAM; 1.
 DR SMART: SM00202; SR; 1.
 DR SMART: SM00020; Tryp. Spec; 1.
 DR PROSITE: PS01180; CUB; 1.
 DR PROSITE: PS01209; LDLa; 1.
 DR PROSITE: PS00068; LDLa_2; 1.
 DR PROSITE: PS00740; MAM_1; UNKNOWN_1.
 DR PROSITE: PS00600; MAM_2; 1.
 DR PROSITE: PS0287; SRCR_2; 1.
 DR PROSITE: PS0240; TRYPSIN_DOM; 1.
 DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.
 DR PROSITE: PS00135; TRYPSIN_SER; 1.
 DR GlycoProtein: Hydrolase; Kinase; Protease; Serine protease.
 DR NON_TER 1
 SC SEQUENCE 777 AA; 87314 MW; A18E2F4ECF06D3A8 CRC64;
 Query Match 18.2%; Score 688; DB 2; Length 777;
 Best Local Similarity 29.3%; Pred. No. 6.2e-40;
 Matches 186; Conservative 93; Mismatches 220; Indels 136; Gaps 22;
 QY 86 LTFSSFDLASCDEGSDLVVYNTLSPEPHALVOLCTYPPSNLTHTSSQNVLLTLI 145
 DB 231 VVFAAFRRNGSTALDIDSLTNGICSGSPYEPTLVTPPP----- 272
 QY 146 TINTERRHGFATFATFQOLPRMSSCG--RLRAQGFNSPYGHPYPIIDCTNIEVNN 203
 DB 273 -----ELP--TDCCGFELMEPNSTFSSPNPDKYFNQASCTIMILNARQ 315
 QY 204 QHVAVRPFYLLBPRACGTCPRDYVEI--NGE-----KYGERSQFVNTSNNKIT 254
 DB 316 KNIGLHGFEDLEN-----INDVVEVRDGEFDSLALLVYGPVVKDLFTYVMT 367
 QY 255 VRPHSDQSYDTGFLAEYLS---YDSDPC--PGQFTCTGTCIRKEFLCDMACTDHS 310
 DB 368 VIFFTNMRTRKRGKAFNTSGYVIGIPPCDDDFQCDGNCIPLGMLCDSPYPCRDSD 427
 QY 311 ELNC-----SCDAGHQFTCKKKFKCPLFMVCDSVNDGDNDSBQSCSPAOTFRCSN 362
 DB 428 EASCVRFLNGTRNSNGLVQFINHS-----IWHI-----ACAEV-----WTTQISN 467
 QY 363 GKCLSKSQCCNGKDCDGDSDSDEASCP-----KVVVVTCTKTYRCINGLCLSKGN 412
 DB 468 EVC-----HLLGLGSANSSWPISTSGGPFVRN-----QAVNGSLILTPS 508
 QY 413 PECGCKE---DCSDGSEKDCDGLSFTROA--RVGCTDABEGEFPVOYSLHALGCG 466
 DB 509 LQCCQSDSLILLCQNHKS-----CGEKVTKVSPKTYGGSDAQAGAPFWVALYHRRS 562
 QY 467 H---ICGASLISPMVLVAACHYIDDRFRYSDPQWAFGLHDSQGRSAPGVQERLTK 523
 DB 563 TDRLICGASLVSQDLVLSAACHCYR---RNLDPKRAVLGLHMQSNLITSPQVRRVVD 618
 QY 524 RIISHPPENDTFVDIALLELEKPAEYSSWVRPICLPDASHVPAGKAIWVGWGHQY 583
 DB 619 QIVNPHYDRARKVNDIMHLEFKNVYTDVIOPICLPEBNQIFIPGRTCSIAAGGYDKI 678
 QY 584 GGTGALLIQKEIRIYINOTTCENLIPQ-QITPRMKCVFLSGVDSCGSDGSPISSTVEA 642
 DB 679 NGSTVDYVLEKADVPVLISNEKCOQOLPBYNLTESMTCAGYEBGGIDSCGSDGSPIMCOE- 737
 QY 643 DGRIFQAGVSMGDCACQANRKGVTTRLPFRDWI 677

DB 738 NNRWFLVGTSPGVQCALPNHGVYRVVSQFIEMI 772
 RESULT 27
 ID ENTK HUMAN STANDARD; PRT; 1019 AA.
 AC P96073;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Enteroleptidase precursor (EC 3.4.21.9) (Enterokinase).
 GN Name=PRSS7; Synonyms=ENTK;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=Duodenum;
 RX MEDLINE=95234679; PubMed=7718557;
 RA Kitamoto Y., Veille R.A., Donis-Keller H., Sadler J.E.;
 RT "cDNA sequence and chromosomal localization of human enterokinase, the
 RT proteolytic activator of trypsinogen."
 RL Biochemistry 34:4562-4568(1995).
 RN (2)
 RP SEQUENCE FROM N.A., AND DISEASE.
 RX MEDLINE=21606074; PubMed=11719902;
 RA Holzinger A., Maier E.M., Buck C., Mayerhofer P.U., Kappler M.,
 RA Haworth J.C., Moroz S.P., Hadorn H.-B., Sadler J.E., Roscher A.A.;
 RT "Mutations in the proenteropeptidase gene are the molecular cause of
 RT congenital enteropeptidase deficiency."
 RL Am. J. Hum. Genet. 70:20-25(2002).
 RN (3)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20289799; PubMed=10830953;
 RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
 RA Park H.-S., Toyoda A., Ishii K., Tokoki Y., Choi D.-K., Groner Y.,
 RA Soeda E., Ohtsuki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K.,
 RA Polley A., Menzel U., Delabar J., Kumpf K., Lehmann R., Paterson D.,
 RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
 RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
 RA Shenhui A., Sasaki T., Nagamine K., Mitsuyama S., Antocarakis S.E.,
 RA Moshima S., Shimizu N., Nordstiek G., Hornisiek K., Brandt P.,
 RA Scharte M., Schen O., Desario A., Reichelt J., Kauter G., Bloeker H.,
 RA Kamber J., Beck A., Klages S., Hennig S., Rieseemann L., Dagand E.,
 RA Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
 RA Lehrach H., Reinhardt R., Yaepp M.-L.;
 RT "The DNA sequence of human chromosome 21."
 RL Nature 405:311-319(2000).
 RN (4)
 RP SEQUENCE OF 749-1019 FROM N.A.
 RC TISSUE=Duodenum;
 RX MEDLINE=94329561; PubMed=8052624;
 RA Kitamoto Y., Yuan X., Wu Q., McCort D.W., Sadler J.E.;
 RT "Enterokinase, the initiator of intestinal digestion, is a mosaic
 RT protease composed of a distinctive assortment of domains."
 RL Proc. Natl. Acad. Sci. U.S.A. 91:7588-7592(1994).
 CC -1- FUNCTION: Responsible for initiating activation of pancreatic
 CC proteolytic proenzymes (trypsin, chymotrypsin and carboxypeptidase
 CC A). It catalyzes the conversion of trypsinogen to trypsin which in
 CC turn activates other proenzymes including chymotrypsinogen,
 CC procarboxypeptidases, and proelastases.
 CC -1- CATALYTIC ACTIVITY: Selective cleavage of 6-Lys-|-Ile-7 bond in
 CC trypsinogen.
 CC -1- SUBUNIT: Heterodimer of a catalytic (light) chain and a
 CC multidomain (heavy) chain linked by a disulfide bond.
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
 CC -1- TISSUE SPECIFICITY: Intestinal brush border.
 CC -1- PTM: The chains are derived from a single precursor that is
 CC cleaved by a trypsin-like protease.
 CC -1- DISEASE: Defects in PRSS7 are a cause of enterokinase deficiency
 CC (MIM:226200); a life-threatening intestinal malabsorption disorder
 CC characterized by diarrhea and failure to thrive.

Qy	323	TCKN----	KFKCLPFWCDSVNDGDNDDEBGCCEPATFFRCNSCKLSKSGQCGNKDCC	378
Db	684	TTNNNGLVFRFIQSIW---- <td>HTACAEK-----WTQISNDVC-----QLL</td> <td>719</td>	HTACAEK-----WTQISNDVC-----QLL	719
Qy	379	GDGSDDEASCPRNVAVVCTKHTYRCNLGCLSKNGPECG-----	KEDCSDGS-----D	426
Db	720	GLSGGNSSKPFFST-----	DGFPFKYKNTAPRPGHILTRPSQCLDDSLRLQCN	768
Qy	427	EKDCDCGLRSFTFRQARVVGGTDADEGEWPMQVSLHALQCGHICGASLISPMVLVSAAHCY		486
Db	769	HKSGCGKTLAAGDIPKIVGSGNAKEGAPWVVGILY-YGGRLLCGASLVSSDWLVSAHCY		827
Qy	487	IDDGFRISDPTQWTAFLGLHDOSORSAFVQERLRKRIISHPPNDPTFYDIALLE		546
Db	828	YG----RULFESKWTALGLHMKNSLTPQVFPRLDEIVINPHYNRKKNDDIAMMLE		883
Qy	547	KPAEYSSWVRICLPDASHVPEPAGKAIWVTGMHQVCGTALILQKEIRVINGTCEN		606
Db	884	FKVNTYDITQPICLPEENQVFPFGNCSIAQGVVYVGGTANILQEDADVLSNERCQ		943
Qy	607	LLPO-QITPRAMVCVFLSGGVDSCQGSQSGPLSSVEADGRIFQAGVSWGDCGAKRNKPG		665
Db	944	QMEPYNTIEMNICAGYEBGGIDSCQGSQGLMCE-NNRFPFLAGVTSFGYKCALPNRPG		1002
Qy	666	VYTRLPLEFRDNK	678	
Db	1003	VYARVSRFTWIIQ	1015	
RESULT 28				
ID	ENTK_MOUSE	STANDARD,	PRT,	1069 AA.
AC	P97435;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	05-JUL-2004 (Rel. 44, Last annotation update)			
DE	Enteropeptidase (EC 3.4.21.9) (Enterokinase).			
GN	Name=Prs87; Synonyma=Entk;			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6; TISSUE=Duodenum;			
RX	MEDLINE=98147142; PubMed=9486188;			
RA	Yuan X., Zheng X., Lu D., Rubin D.C., Pung C.Y.M., Sadler J.E.;			
RT	"Structure of murine enterokinase (enteropeptidase) and expression in			
RT	small intestine during development."			
RL	Am. J. Physiol. 274:G342-G349(1998).			
CC	-1- FUNCTION: Responsible for initiating activation of pancreatic			
CC	proteolytic proenzymes (trypsin, chymotrypsin and carboxypeptidase			
CC	A). It catalyzes the conversion of trypsinogen to trypsin which in			
CC	turn activates other proenzymes including chymotrypsinogen,			
CC	procarboxypeptidases, and proelastases (By similarity).			
CC	-1- CATALYTIC ACTIVITY: Selective cleavage of 6-Lys-Ile-7 bond in			
CC	trypsinogen.			
CC	-1- SUBUNIT: Heterodimer of a catalytic (light) chain and a			
CC	multidomain (heavy) chain linked by a disulfide bond (By			
CC	similarity).			
CC	-1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).			
CC	-1- PM: The chains are derived from a single precursor that is			
CC	cleaved by a trypsin-like protease (By similarity).			
CC	-1- SIMILARITY: Belongs to peptidase family S1.			
CC	-1- SIMILARITY: Contains 2 CUB domains.			
CC	-1- SIMILARITY: Contains 2 LDI-receptor class A domains.			
CC	-1- SIMILARITY: Contains 1 MAM domain.			
CC	-1- SIMILARITY: Contains 1 SEA domain.			
CC	-1- SIMILARITY: Contains 1 SRCR domain.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on items			

CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/1sb-sib.ch).	
CC	or send an email to license@1sb-sib.ch .	
CC	-----	
DR	EMBL; U73378; AAB37317.1; -.	
DR	HSSP; P98072; 1KB.	
DR	MEROPS; S01.156; -.	
DR	MGD; MGI:1197523; Prsa7.	
DR	InterPro; IPR000859; CUB.	
DR	InterPro; IPR002172; LDL_receptor_A.	
DR	InterPro; IPR000998; MAM.	
DR	InterPro; IPR001254; Peptidase_S1.	
DR	InterPro; IPR001314; Peptidase_S1A.	
DR	InterPro; IPR011163; Pept_S1A_enterop.	
DR	InterPro; IPR009003; Pept_Ser_Cys.	
DR	InterPro; IPR000082; SEA.	
DR	InterPro; IPR001190; Srcr_receptor.	
DR	Pfam; PF00431; CUB; 2.	
DR	Pfam; PF00057; Ldl_recept_a; 2.	
DR	Pfam; PF00629; MAM; 1.	
DR	Pfam; PF01390; SEA; 1.	
DR	Pfam; PF00530; SRCR; 1.	
DR	Pfam; PF00089; Trypsin; 1.	
DR	PIRSP; PIRSF001138; Enteropeptidase; 1.	
DR	PRINTS; PR00722; CHYMOTRYPSIN.	
DR	PRINTS; PR00261; LDLRECEPTOR.	
DR	PRINTS; PR00020; MAMDOMAIN.	
DR	SMART; SM00042; CUB; 2.	
DR	SMART; SM00192; LDla; 2.	
DR	SMART; SM00137; MAM; 1.	
DR	SMART; SM00200; SEA; 1.	
DR	SMART; SM00202; SR; 1.	
DR	SMART; SM00020; Tryp_Spc; 1.	
DR	PROSITE; PS01180; CUB; 2.	
DR	PROSITE; PS01209; LDLRA_1; 2.	
DR	PROSITE; PS50068; LDLRA_2; 2.	
DR	PROSITE; PS00740; MAM_1; 1.	
DR	PROSITE; PS50060; MAM_2; 1.	
DR	PROSITE; PS50024; SEA; 1.	
DR	PROSITE; PS500420; SRCR_1; FALSE_NEG.	
DR	PROSITE; PS50287; SRCR_2; 1.	
DR	PROSITE; PS50240; TRYPsin_DOM; 1.	
DR	PROSITE; PS00134; TRYPsin_HIS; 1.	
DR	PROSITE; PS00135; TRYPsin_SER; 1.	
KW	Glycoprotein; Hydrolase; Repeat; Serine protease; Signal-anchor; Transmembrane; Zymogen.	
KW	CHAIN 1 829	
FT	CHAIN 830 1069	
FT	DOMAIN 1 18	
FT	TRANSMEM 19 47	
FT		
FT	DOMAIN 48 1069	
FT	DOMAIN 52 169	
FT	DOMAIN 227 268	
FT	DOMAIN 270 379	
FT	DOMAIN 387 549	
FT	DOMAIN 569 679	
FT	DOMAIN 686 724	
FT	DOMAIN 723 816	
FT	DOMAIN 830 1069	
FT	ACT_SITE 874 874	
FT	ACT_SITE 925 925	
FT	ACT_SITE 1021 1021	
FT	DISULFID 229 242	
FT	DISULFID 236 255	
FT	DISULFID 249 266	
FT	DISULFID 688 700	
FT	DISULFID 695 713	
FT	DISULFID 707 722	
FT	DISULFID 817 945	
FT	DISULFID 859 875	
FT	DISULFID 959 1027	
FT		
FT	DOMAIN 48 1069	
FT	DOMAIN 52 169	
FT	DOMAIN 227 268	
FT	DOMAIN 270 379	
FT	DOMAIN 387 549	
FT	DOMAIN 569 679	
FT	DOMAIN 686 724	
FT	DOMAIN 723 816	
FT	DOMAIN 830 1069	
FT	ACT_SITE 874 874	
FT	ACT_SITE 925 925	
FT	ACT_SITE 1021 1021	
FT	DISULFID 229 242	
FT	DISULFID 236 255	
FT	DISULFID 249 266	
FT	DISULFID 688 700	
FT	DISULFID 695 713	
FT	DISULFID 707 722	
FT	DISULFID 817 945	
FT	DISULFID 859 875	
FT	DISULFID 959 1027	
FT		
FT	DOMAIN 48 1069	
FT	DOMAIN 52 169	
FT	DOMAIN 227 268	
FT	DOMAIN 270 379	
FT	DOMAIN 387 549	
FT	DOMAIN 569 679	
FT	DOMAIN 686 724	
FT	DOMAIN 723 816	
FT	DOMAIN 830 1069	
FT	ACT_SITE 874 874	
FT	ACT_SITE 925 925	
FT	ACT_SITE 1021 1021	
FT	DISULFID 229 242	
FT	DISULFID 236 255	
FT	DISULFID 249 266	
FT	DISULFID 688 700	
FT	DISULFID 695 713	
FT	DISULFID 707 722	
FT	DISULFID 817 945	
FT	DISULFID 859 875	
FT	DISULFID 959 1027	
FT		
FT	DOMAIN 48 1069	
FT	DOMAIN 52 169	
FT	DOMAIN 227 268	
FT	DOMAIN 270 379	
FT	DOMAIN 387 549	

FT DISULFID 991 1006 By similarity.
 FT DISULFID 1017 1045 By similarity.
 FT CARBOHYD 147 147 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 197 197 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 212 212 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 373 373 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 380 380 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 433 433 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 515 515 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 579 579 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 675 675 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 727 727 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 751 751 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 770 770 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 791 791 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 897 897 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 936 936 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 999 999 N-linked (GlcNAc...) (Potential).
 SQ SEQUENCE 1069 AA, 118735 MW, E62549B463743C3D CRC64;

Query Match 17.9%; Score 677.5; DB 1; Length 1069;
 Best Local Similarity 29.2%; Pred. No. 4,9e-39;
 Matches 186; Conservative 93; Mismatches 220; Indels 137; Gaps 23;

QY 86 LTRSFPLASCDERGSGLVTVYNTLSPEPHALVOLCGYPPSYNLTFFHSSQNVLLITLI 145
 DB 522 VVFVAFNRRCSTALDIDSLTNGICSGSPYEPETLVTPPP----- 563
 QY 146 TINTERRRHGFATFPQLPRMSSCGG-RLRKAQSTFSPYRGHYPPNIDTWNIEVNN 203
 DB 564 -----ELRP--TDCGPELWEPNSTFSSPMPDXYPAQASCIMWLNARQ 606
 QY 204 QHVAVRKFFYLLERPRACGCPDYVEI--NGE-----KYCGESQFVTSNSKIT 254
 DB 607 KNIGLHGEEDLEN-----INDVYEVADGSEFDSLLAVYGPPEVADLFSTTRMT 656
 QY 255 VRHSQSYTDGTGFLAAYLS--YDSSDPC-PGQFTCTRCRCIKELRCGMADCTDHS 310
 DB 659 VIETNNMETRRKGRKAFPTGYYLGIPEPCODDEFQCKDNCLPLGNLCSYPHCRDSD 718
 QY 311 ELNLC-----SCDAHQFTCKKFKFKPLFWVDSVNDGSDNDEGCGCPAQTFRCSN 362
 DB 719 EASCVRFNGTRSNNGLVQFNIHS-----IWHI--ACAEEN-----WTTQISN 758
 QY 363 GKCLSKSQCGKXKDCGSDGDEASCP-----KVNVTCTKHYRCINGLCLSKGN 412
 DB 759 EVC-----HLLGLGSANSMPLSSVCGGFVAVN-----QAPNSGLITLPS 799
 QY 413 PECDGKE---DCSDGSDKDCGSLRFRQA--RVVGTDADGEMPMQVSLHALGQG 466
 DB 800 LQCSQDSLILLOCHNKS-----CGEKRVYQKVPKIVGSDAQAGAMPVVALYHRDS 853
 QY 467 H---IGASLISPMWLSAAHCYIDDGFRYSDDTQMTAFGLHDSQSRASPGQERRLK 523
 DB 854 TDRLLCASLIVSSWMLVSAHCVYR---RNLDTKRTAVLGLHMOGNSLSPQVVRVVD 909
 QY 524 RIISHPEFNDFTPDYALALELEKPAEYSSWVRPICHPDASHVPACKAIVWTGHTQY 583
 DB 910 QIVINHYDRARRKNDIAMMHELEKVAITYIQCIPRENOQIFIPERTCSIAWGWDKI 969
 QY 584 -GGTALLDKGEIRVINQTTCENTLFPQ-QTPRMCMVGLSGVDSQCGDSGGPLSSVE 641
 DB 970 NAGSTVAVLEKADVPRLISNEKCCQQLREYNTSEMICAGYEGGIDSCQDSGGLMCOE 1029
 QY 642 ADGRIPQAGVSWDGAQRNKPGVYTRLLPLFRMI 677
 DB 1030 -NNKWFIVGVTSFGVQCALPNHFGVYRVVSQFIEMI 1064

RESULT 29
 CORI_MOUSE STANDARD; PRT; 1113 AA.
 AC Q92319;

DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Atrial natriuretic peptide-converting enzyme (EC 3.4.21.-) (pro-ANP-
 DE converting enzyme) (Corin) (low density lipoprotein receptor related
 DE protein 4).
 GN Name=Corin; Synonyms=Crm, Lrp4;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98429596; Pubmed=9756624;
 RA Tomita Y., Kim D.-H., Magocsi K., Fujino T., Yamamoto T.T.;
 RT "A novel low-density lipoprotein receptor-related protein with type II
 RT membrane protein-like structure is abundant in heart."
 RL J. Biochem. 124:784-789 (1998).
 CC - FUNCTION: Converts pro-ANP to ANP. Cleaves pro-ANP specifically
 CC between Arg-122 and Ser-123 (By similarity).
 CC - SUBCELLULAR LOCATION: Type II membrane protein.
 CC - TISSUE SPECIFICITY: Highly expressed in heart.
 CC - SIMILARITY: Belongs to peptidase family S1.
 CC - SIMILARITY: Contains 2 Fz-like (Fz) domains.
 CC - SIMILARITY: Contains 7 LDL-receptor class A domains.
 CC - SIMILARITY: Contains 1 SRCR domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AB013874; BAA34371.1; -.
 CC PIR; J03015; J03015.
 CC HSRP; P01130; 1A0J.
 CC MEROPS; S01.019; -.
 CC MGD; MGI:1349451; Corin.
 DR InterPro: IPR000024; Fz domain.
 DR InterPro: IPR002172; LDL_receptor_A.
 DR InterPro: IPR001254; Peptidase_S1.
 DR InterPro: IPR001314; Peptidase_S1A.
 DR InterPro: IPR009003; Pept_Ser_Cys.
 DR InterPro: IPR001190; Srcr_receptor.
 DR Pfam; PF01392; Fz_2.
 DR Pfam; PF00057; Ldl_recept_a; 6.
 DR Pfam; PF00069; Trypsin_1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR PRINTS; PR00261; LDLRECEPTOR.
 DR SMART; SM00063; FRI; 2.
 DR SMART; SM00192; LDLa; 7.
 DR SMART; SM00202; SR; 1.
 DR SMART; SM00201; TYP_SPC; 1.
 DR PROSITE; PS50038; Fz_2.
 DR PROSITE; PS01209; LDLa_1; 6.
 DR PROSITE; PS50068; LDLa_2; 7.
 DR PROSITE; PS00420; SRCR_1; FALSE_NEG.
 DR PROSITE; PS50287; SRCR_2; 1.
 DR PROSITE; PS50240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; FALSE_NEG.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 KW Glycoprotein; Hydrolase; Repeat; Serine protease; Signal-anchor;
 KW Transmembrane.
 FT DOMAIN 1 112 Cytoplasmic (Potential).
 FT TRANSMEM 113 133 Signal-anchor for type II membrane
 FT protein (Potential).
 FT DOMAIN 134 1113 Extracellular (Potential).
 FT DOMAIN 134 327 Fz_1.
 FT DOMAIN 336 372 LDL-receptor class A 1.
 FT DOMAIN 373 408 LDL-receptor class A 2.
 FT DOMAIN 409 445 LDL-receptor class A 3.

```

FT DOMAIN 446 483 LDL-receptor class A 4.
FT DOMAIN 518 641 Fz 2.
FT DOMAIN 647 682 LDL-receptor class A 5.
FT DOMAIN 683 721 LDL-receptor class A 6.
FT DOMAIN 722 757 LDL-receptor class A 7.
FT DOMAIN 758 853 SRCR.
FT ACT_SITE 869 1113 serine protease.
FT ACT_SITE 910 910 Charge relay system (By similarity).
FT ACT_SITE 959 959 Charge relay system (By similarity).
FT ACT_SITE 1052 1052 Charge relay system (By similarity).
FT DISULFID 857 979 By similarity.
FT DISULFID 895 911 By similarity.
FT DISULFID 1022 1037 By similarity.
FT DISULFID 1048 1077 By similarity.
FT CARBOHYD 147 147 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 202 202 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 208 208 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 298 298 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 317 317 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 373 373 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 411 411 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 444 444 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 481 481 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 519 519 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 537 537 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 635 635 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 719 719 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 765 765 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 828 828 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 970 970 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 1089 1089 N-linked (GlcNAc...) (Potential).
SQ SEQUENCE 1113 AA; 122964 MW; B845B2C5F20DDEBC CRC64;

Query Match 17.5%; Score 663.5; DB 1; Length 1113;
Best Local Similarity 32.7%; Pred. No. 4.9e-38;
Matches 158; Conservative 67; Mismatches 187; Indels 71; Gaps 16;

QY 256 RFHSDGYTDGFLAEVLYSDSDPCQFCTCRGCRKRLRDGMADCDHSDDELNCS 315
DB QPESSDNQCTLLP---NEVEECSPHFRCRGRVLSRRCDGQDCDDDEECG 683
QY 316 CDAGHQFTCK-NKFCPLFWVCDSDVNDGSDSDGCS-CPAQTFRCNGKSLKSQOCN 373
DB 684 CKERALMECPNKKQCLKHTLTCDFPRCDSDMERKNSFCQDNLECANHECVRRDLCD 743
QY 374 GKDDCGSDGSDASCPRKN-----VTCYKTYRCLNG-----LC----- 407
DB 744 GWDVCSDDSDMGCVTLSSKNGSSSLTVHKSAXEHVCAQWRETLSQLACKQMGLEGP 803
QY 408 -LSKGNBECQKE-----DCSDGSDKDP-----C---DCGLSFT 438
DB 804 SVTLKILPQEQQWLRLYPNMENLNGSTLQELVYRHSFSPRSISLSCSKQDGRPPA 863
QY 439 R-QARVVGSTADAGGEMPMQVSLHALGQHICGASLISPNMLVSAHCYIDRGRYSDF 497
DB 864 RMNRITLIGRRSRGRPMQCSLOSBSGHICGVLLAKKVVLYAHCP-BGR-----EDA 918
QY 498 TQMTAFGLDQSGRSPAGVGERLKRILISHPFNDFTFDVIALLELEKAEVSSMVR 557
DB 919 DVMKVVFGINNLDHPSG-FMQTRFVKTLIPRYSRAVVDYDISVELSDINETSYP 977
QY 558 ICLPDASHVFPAGKAIWVGHWGTOYCGTALIIQKGRIRINOTGENTLLPQC-IIPRM 616
DB 978 VCLPSPEBYLEPTDYCYITGMH--MGNKMPFKLQEGSVRIILPEQCSYDMKTITRM 1035
QY 617 MCVFPLSGVSDSCGSDGSLSVLEADGRIFQAGVSWGDCQAOB-NKPGVYTRLLPFRD 675
DB 1036 ICAIGESGTVDSCMGDSGLVCEPRGQWTLFGILTMSGVSFSLVGLPGVYSNVSTYVG 1095
QY 676 WIK 678
DB 1096 WIE 1098

```

```

RESULT 30
Q9DGR2 PRELIMINARY; PRT; 767 AA.
AC Q9DGR2;
DT 01-MAR-2001 (Tremblrel, 16, Created)
DT 01-MAR-2001 (Tremblrel, 16, Last sequence update)
DT 01-MAR-2004 (Tremblrel, 26, Last annotation update)
DE Embryonic serine protease-2.
GN Name=xsp-2;
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodidae; Xenopus.
NCBI_TaxID=8355;
RX MEDLINE=20363741; PubMed=10903452;
RA Yamada K., Takabatake T., Takeshima K.;
RT "Isolation and characterization of three novel serine protease genes
RT from Xenopus laevis";
RL Gene 252:209-216 (2000).
CC -1- SIMILARITY: Belongs to peptidase family S1.
DR EMBL; AB038497; BAB08217.1; -.
DR HSSP; P00760; IEZX.
DR MEROPS; S01.049; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0005044; F:scavenger receptor activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR InterPro; IPR009003; Pept_Ser_Cys.
DR Pfam; PF000190; Serc_receptor.
DR Pfam; PF00089; Ldl_recept_a; 4.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00261; LDLRECEPTOR.
DR SMART; SM00192; LDLa; 8.
DR SMART; SM00020; TRYR_SPC; 1.
DR PROSITE; PS001209; LDLA_1; 8.
DR PROSITE; PS50068; LDLA_2; 2.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolyase; Protease; Serine protease.
SQ SEQUENCE 767 AA; 86001 MW; E056A38796DE96E CRC64;

Query Match 17.0%; Score 644.5; DB 2; Length 767;
Best Local Similarity 26.6%; Pred. No. 7e-37;
Matches 170; Conservative 72; Mismatches 214; Indels 183; Gaps 17;

QY 193 DCTN-----IEVNNQVKKRKFYLLERRACGTPKQVLENGSKYC-GERSQ 243
DB 150 DCPYGDERNCAKTPPTPCQWCSYY-----TCYVQICNVQDCPYGDER 200
QY 244 FVVTNSNKTIVFPHSQSYTDGFLAEVLYSDSDPCQFCTCRG----- 290
DB 201 NCATKTSIPTQWYC-STSTSCIYGYQ-----CNGVQCAVDDERNCAKTPSI 251
QY 291 -----RCIRKELRCDGADCTHSDDELNCSGDAHQFTCK-----NKFCKP 331
DB 252 PTCQLCYSYTYTCTIAYQICNGVLDPCFVDDERNCAVATSTPTCQICYWDFMEDYTCIY 311
QY 332 LFWVCDSDVNDGSDSDGSC-----SCPAQFRFSN-----CKLSKSGQCNCKD 376
DB 312 AYQWCDGVRQCYGDDERNCAVATTTATTSTPTCQIYCMNPMYYTCTIAYQWCDGVR 371
QY 377 DCGDSDGSDASC-PRVNVVTCYKTYRCLNGCLSKNGPECCKEDCSGSDKDC----- 430

```

```

Db      372 QCYGDBELNCDRTTAYCEK---RCSSVSCVLSQWCDGVSDCPYGEDMSCVSLYP 428
QY      431 -----430
Db      429 ADFOLOVYSTSVSAMLFCSDYMNDFGRFACODFGYNGSSYNRYDTLMSPYAPNGYFKL 488
QY      431 -----DCGLRSFTROARVVGSTDADEGEMWQVSL 460
Db      489 YSGYWRSEKFTSVQYSSYCSGNVSLHCTSCGVSNNSLVSRIVGTFANLGNMFWQVNL 548
QY      461 HALGQGHICGASLISPMWLVSAHCTYDDKGFYSPTQWTAFLGLHDQSORSAPOVER 520
Db      549 QYI-TGVLGSGSIIISPMIVYAAHCVYGS---YSSASGWRVFAG--TLTKPSYYNASAY 601
QY      521 RLKRIISHPPFNDFTFYDIALLELEKPAEYSSMWRPCLPDASHVPFAGKAIWVTGMH 580
Db      602 FVERIIVHPGKSYTYNDIALMKLRDEITFGYTTQVPCLPNSGMFWBAGTTTWISGMS 661
QY      581 TOYGGTALILQKEIRVINQTCEN--LIPQITPRMVCVFLSGGVDSQGDGSGPLS 638
Db      662 TYEGGSYSTYLQYVAIPLIDSNVNCQSYVYNGQITSSMICAGYLSGGVDTQGDGSGPLV 721
QY      639 SVEADGRIFOAGVVSQDGCQQRNKEGVYTRLPFLFRDVI 677
Db      722 N-KRNGTWMLVGDTSGDGCARANKPGVYGNVTFLEMI 759

```

Search completed: November 29, 2004, 08:32:23
 Job time : 123.685 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 29, 2004, 08:19:58 ; Search time 25.3127 Seconds
(without alignments)
2596.165 Million cell updates/sec

Title: US-09-936-333-5
Perfect score: 3789
Sequence: 1 MAERVYMLPPRARSLKSFV.....PGVYTRLPFRMDIKENTGV 683

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3197	84.4	855	2 JCT7721	membrane-bound arg
2	718.5	19.0	1034	1 A53663	enteropeptidase (E
3	703.5	18.6	1035	1 A43090	enteropeptidase (E
4	683.5	18.0	1019	1 A56318	enteropeptidase (E
5	663.5	17.5	1113	2 JE0315	low-density lipopr
6	576.5	15.2	1524	2 T30337	polypeptide - Afri
7	565	14.9	699	1 I54763	Ra-reactive factor
8	545.5	14.4	638	1 KQMSPL	plasma kallikrein
9	526.5	13.9	638	1 KQRTPL	plasma kallikrein
10	506	13.4	613	2 S15468	complement C3b/C4b
11	504	13.3	790	1 B1545	plasmin (EC 3.4.21
12	503	13.3	460	2 B61545	plasmin (EC 3.4.21
13	501.5	13.2	786	1 A47547	serine proteinase
14	499	13.0	638	1 KQHUP	plasma kallikrein
15	492.5	13.0	583	2 A29154	complement factor
16	492.5	13.0	810	1 PLHU	plasmin (EC 3.4.21
17	491.5	13.0	812	1 PLMS	plasmin (EC 3.4.21
18	491	13.0	416	1 KFRBO	coagulation factor
19	490	12.9	417	1 S00845	hepsin (EC 3.4.21.
20	484.5	12.8	343	1 A57014	protease (EC 3.4.
21	484.5	12.8	810	2 KFHU	coagulation factor
22	483	12.7	810	2 I46260	plasmin (EC 3.4.21
23	481.5	12.7	812	1 PLBO	plasmin (EC 3.4.21
24	481	12.7	810	2 B30848	plasmin (EC 3.4.21
25	480.5	12.7	416	1 S33777	hepsin (EC 3.4.21.
26	480	12.7	625	1 KFHU1	coagulation factor
27	474.5	12.5	452	1 A50351	Ra-reactive factor
28	474	12.5	686	1 A59271	coagulation factor
29	472	12.5	275	2 S40007	trypsin (EC 3.4.21

30	472	12.5	455	2 A61545	plasmin (EC 3.4.21
31	468.5	12.4	2616	2 A57096	nucl protein prec
32	467	12.3	275	2 A50005	trypsin (EC 3.4.21
33	467	12.3	562	1 UKHTT	t-plasminogen acti
34	462	12.2	274	2 S35339	trypsin (EC 3.4.21
35	461.5	12.2	267	2 S40006	trypsin (EC 3.4.21
36	460	12.1	1420	2 A32869	apolipoprotein(a)
37	452.5	11.9	560	1 JC4795	plasma hyaluronan
38	451	11.9	238	1 TRWVSJ	trypsin-like prote
39	449.5	11.9	4548	1 S00557	apoptotain(a) (EC
40	442.5	11.7	1004	2 T30338	oviductin (EC 3.4.
41	441	11.6	254	1 TRWV3Y	trypsin-like prote
42	439	11.6	270	2 S56160	maet cell trypsinase
43	438.5	11.6	459	2 JQ0419	coagulation factor
44	437.5	11.5	247	1 TRDC	trypsin (EC 3.4.21
45	436	11.5	242	2 S49489	trypsin (EC 3.4.21
46	436	11.5	263	2 A21195	chymotrypsin (EC 3
47	434	11.5	263	1 KYRTB	chymotrypsin (EC 3
48	434	11.5	265	2 T15451	hypothetical prote
49	433.5	11.4	273	2 A47246	trypsinase (EC 3.4.2
50	433	11.4	273	2 S35340	trypsin (EC 3.4.21
51	431.5	11.4	221	1 TRPGTR	trypsin (EC 3.4.21
52	431	11.4	276	2 A38654	maet cell proteina
53	428.5	11.3	247	1 B25852	trypsin (EC 3.4.21
54	427.5	11.3	247	2 A27547	trypsin (EC 3.4.21
55	427.5	11.3	622	1 TBHU	thrombin (EC 3.4.2
56	426.5	11.3	407	1 KFRB07	coagulation factor
57	426	11.2	231	2 S31778	trypsin (EC 3.4.21
58	425.5	11.2	558	2 JC5878	plasma hyaluronan
59	425	11.2	275	2 A32410	trypsinase (EC 3.4.2
60	424	11.2	274	2 JC4171	chymotrypsin (EC 3
61	422	11.1	263	2 A31299	chymotrypsin (EC 3
62	422	11.1	275	2 B35863	trypsinase (EC 3.4.2
63	421	11.1	246	2 B25528	trypsin (EC 3.4.21
64	421	11.1	246	2 JQ1472	trypsin (EC 3.4.21
65	421	11.1	275	2 A35863	trypsinase (EC 3.4.2
66	420.5	11.1	266	2 S54146	trypsin (EC 3.4.21
67	420.5	11.1	304	2 S33496	trypsin (EC 3.4.21
68	420.5	11.1	559	1 A35029	t-plasminogen acti
69	420.5	11.1	705	1 C1HURB	complement subcomp
70	420	11.1	245	1 KYBOB	chymotrypsin (EC 3
71	419.5	11.1	559	1 A29941	t-plasminogen acti
72	419.5	11.1	761	2 JC5759	brain-specific ser
73	419	11.1	246	2 JQ1471	trypsin (EC 3.4.21
74	418.5	11.0	456	1 KXBO	prolein C (activat
75	418	11.0	229	1 TRDPS	trypsin (EC 3.4.21
76	418	11.0	618	2 A35827	thrombin (EC 3.4.2
77	417.5	11.0	246	1 TRRT1	trypsin (EC 3.4.21
78	417	11.0	242	2 S31776	trypsin (EC 3.4.21
79	417	11.0	242	2 S31775	trypsin (EC 3.4.21
80	417	11.0	275	2 C35863	trypsinase (EC 3.4.2
81	417	11.0	462	1 EXRT	coagulation factor
82	415.5	11.0	247	2 S13813	trypsin (EC 3.4.21
83	415	11.0	247	2 S05494	trypsin (EC 3.4.21
84	415	11.0	432	1 S18932	u-plasminogen acti
85	413.5	10.9	274	2 A45754	trypsinase (EC 3.4.2
86	412.5	10.9	238	2 S31779	trypsin (EC 3.4.21
87	412.5	10.9	238	2 S47538	trypsin (EC 3.4.21
88	412	10.9	263	2 S47537	chymotrypsin (EC 3
89	411.5	10.9	250	2 S55493	serine proteinase
90	411	10.8	247	1 A25852	trypsin (EC 3.4.21
91	410	10.8	266	2 JC4850	trypsin-like prote
92	409.5	10.8	237	1 TRCV1	trypsin (EC 3.4.21
93	409.5	10.8	274	2 I47078	coagulation factor
94	408	10.8	245	1 KYBOA	chymotrypsin (EC 3
95	408	10.8	259	2 I38363	trypsin (EC 3.4.21
96	407.5	10.8	264	2 I38136	chymotrypsin-like
97	407	10.7	492	1 EXBO	coagulation factor
98	406.5	10.7	246	1 TRDGC	trypsin (EC 3.4.21
99	406.5	10.7	466	1 KFHU7	coagulation factor
100	406.5	10.7	617	2 S10511	thrombin (EC 3.4.2

plasmin (EC 3.4.21
nucl protein prec
trypsin (EC 3.4.21
t-plasminogen acti
trypsin (EC 3.4.21
trypsin (EC 3.4.21
apolipoprotein(a)
plasma hyaluronan
trypsin-like prote
apoptotain(a) (EC
oviductin (EC 3.4.
trypsin-like prote
maet cell trypsinase
coagulation factor
trypsin (EC 3.4.21
trypsin (EC 3.4.21
chymotrypsin (EC 3
chymotrypsin (EC 3
hypothetical prote
trypsinase (EC 3.4.2
trypsin (EC 3.4.21
trypsin (EC 3.4.21
maet cell proteina
trypsin (EC 3.4.21
trypsin (EC 3.4.21
thrombin (EC 3.4.2
coagulation factor
trypsin (EC 3.4.21
plasma hyaluronan
trypsinase (EC 3.4.2
chymotrypsin (EC 3
chymotrypsin (EC 3
trypsin (EC 3.4.21
chymotrypsin-like
coagulation factor
trypsin (EC 3.4.21
coagulation factor
thrombin (EC 3.4.2

ALIGNMENTS

RESULT 1

JC7731

membrane-bound arginine-specific serine proteinase precursor - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004

C:Accession: JC7731; JC7775

R:Kishi, K.; Yamazaki, K.; Yasuda, I.; Yahagi, N.; Ichinose, M.; Tsuchiya, Y.; Achauda, J. Biochem. 130, 425-430, 2001

A:Title: Characterization of a membrane-bound arginine-specific serine protease from rat

A:Reference number: JC7731; MUID:21421307; PMID:11530019

A:Accession: JC7731

A:Molecule type: mRNA

A:Residues: 1-855 <KTS>

A:Cross-references: UNIPROT:Q9JUI7; DDBJ:AB049189

A:Experimental source: strain Male, 7-week-old

R:Satomi, S.; Yamazaki, Y.; Tsuchi, S.; Hitomi, Y.; Iwanaga, T.; Fushiki, T.

Biochem. Biophys. Res. Commun. 287, 995-1002, 2001

A:Title: A role for membrane-type serine protease (MT-SPL) in intestinal epithelial turn

A:Reference number: JC7775; PMID:11573963

A:Contents: Small intestine

A:Accession: JC7775

A:Molecule type: mRNA

A:Residues: 1-855 <SAT>

A:Cross-references: DDBJ:AB037898

C:Comment: This enzyme, an epithelial-derived, type II integral membrane serine protease of specific proteins or peptides on the brushborder membranes. It also participates in

lial migration and/or cell loss.

C:Genetics:

A:Gene: mt-spl

A:Map position: basolateral cell surface

C:Superfamily: membrane-bound arginine-specific serine proteinase

C:Keywords: protein digestion

Query Match 84.4%; Score 3197; DB 2; Length 855;

Best Local Similarity 83.0%; Pred. No. 3.7e-203;

Matches 567; Conservative 59; Mismatches 57; Indels 0; Gaps 0;

```
QY 1 MAERVVMLPPRAASLSPVVTSVARPTDSKYQRTQDSSCSGLHARGVELMRFTTPG 60
DB 173 MAVRRVVTLPRAALSKSPVLTVAAPDIPRMQKRDNSCSFALHARGVTRFTTPG 232
QY 61 PPDSPYAHARCOMALRGDADSVLSTFRSPDLASCERGSDDLVTYVNTLSMPHFALVQ 120
DB 233 FPNSPYAHARCOMALRGDADSVLSTFRSPDLASCERGSDDLVTYVNTLSMPHFALVQ 292
QY 121 LCGTYPSPYNLTFRHSONVLLITLITERRHHPGFATFFQLPRMSSCGRLRKAQGTEN 180
DB 293 LCGTSPSPYNLTFRHSONVLLITLITERRHHPGFATFFQLPRMSSCGRLRKAQGTEN 352
QY 181 SPYRPGHYPRINDCTNMIIEVPPNNOHVVRKRFLLPRACGCPDYVEINEKTCGE 240
DB 353 SPYRPGHYPRINDCTNMIIEVPPNNOHVVRKRFLLPRACGCPDYVEINEKTCGE 412
QY 241 RSQFVTVSNKTIIVRFRHSDOSYDTGTFLAEYLSYDSDPCPOGFTCRGCIKELRCD 300
DB 413 RSQFVTVSNKTIIVRFRHSDOSYDTGTGTFLAEYLSYDSDPCPOGFTCRGCIKELRCD 472
QY 301 GMADCTHSDDELNCSDAHQFTCKNFKFPLFVWCDSDVNDGSDNDEGSCSPAQTFRC 360
DB 473 GMADCPYSDERHRCNATQHFQCKNOFKFPLFVWCDSDVNDGSDNDEGSCSPAQTFRC 532
QY 361 SNGKCLSKSQOQCNKDCGSGSDASCPKVVVCTGTRCLNGLCLSKGNPECDKED 420
DB 533 SNGKCLPQSQOQCNKDCGSGSDASCDVNAVASCTTYRCONGLCLNKNPECDKED 592
QY 421 CSDSDERKDCDGLRFTQARVVVGTDADGEPWPMOVLHLAQGHICGASLSPMLV 480
DB 593 CSDSDERKDCDGLRFTQARVVVGTDADGEPWPMOVLHLAQGHICGASLSPMLV 652
QY 481 SAACHYIDRGRFYSDEPTQWTAFLGLHDQSRASAPGVQERRLKRILSHPFNDFTFDYDI 540
```

RESULT 2

A53663

enteropeptidase (EC 3.4.21.9) precursor [validated] - pig

N:Alternate names: enterokinase

C:Species: Sus scrofa domestica (domestic pig)

C:Date: 07-Oct-1994 #sequence_revision 09-Aug-1996 #text_change 28-Apr-2003

C:Accession: A53663

R:Matsumura, M.; Ichinose, M.; Yahagi, N.; Kakei, N.; Tsukada, S.; Miki, K.; Kurokawa, J. Biol. Chem. 269, 19976-19982, 1994

A:Title: Structural characterization of porcine enteropeptidase.

A:Reference number: A53663; MUID:94327548; PMID:8051081

A:Accession: A53663

A:Molecule type: mRNA

A:Residues: 1-1034 <MAT>

A:Cross-references: GB:D30799; NID:9505122; PIDN:BA06459.1; PID:9505123

A:Note: parts of this sequence, including the amino ends of three chains isolated from ct

C:Comment: The mechanism of association with the membrane of the intestinal brush border

C:Complex: Mature enteropeptidase is variously reported to contain two (heavy and light)

ed by a disulfide bond. Possibly, conversion from membrane-bound to soluble forms involve

C:Function:

A:Description: cleaves activation peptide from trypsinogen to produce active trypsin

A:Pathway: intestinal digestive hydrolase cascade

C:Superfamily: enteropeptidase; Clr/Cls repeat homology; LDL receptor ligand-binding rep

C:Keywords: glycoprotein; hydrolase; serine proteinase; transmembrane protein; zymogen

F:12-33/Domain: transmembrane #status predicted <TM>

F:52-117/Product: enteropeptidase mini chain #status predicted <MCH>

F:118-799/Product: enteropeptidase heavy chain #status predicted <HCH>

F:199-239/Domain: LDL receptor ligand-binding repeat homology <LDL1>

F:357-516/Domain: MAM homology <MAM>

F:511-646/Domain: Clr/Cls repeat homology <Clr>

F:658-692/Domain: LDL receptor ligand-binding repeat homology <LDL2>

F:693-798/Domain: scavenger receptor cysteine-rich domain homology #status atypical <SRC

F:800-1034/Product: enteropeptidase light chain #status predicted <LCH>

F:116-147/Domain: trypsin homology <TRY>

F:787-911, 825-841, 925-992, 956-971, 982-1010/Disulfide bonds: #status predicted

F:840-891, 986/Active site: His, Asp, Ser #status predicted

Query Match 19.0%; Score 718.5; DB 1; Length 1034;

Best Local Similarity 32.4%; Pred. No. 1e-39;

Matches 181; Conservative 87; Mismatches 189; Indels 101; Gaps 22;

```
QY 161 QLFPMSSCGG--RLRAQGTSPSPYRPGHYPRINDCTNMIIEVPPNNOHVVRKRFLLPR 218
DB 534 ELP--TDCCGFPELMPEPNTFTSMNPNNYPNOACFVNMIMNQKGNIDLHPEEPDEN- 590
QY 219 RRAQGTCPDYVEI--NGEK-----YGERGQFVWTSNSKTIIVRFRHSDOSYDTGTFL 269
DB 591 -----LADVEIRGSEEDSLILAAYVTPGPEVEYVFTTNMTVLFTINDALTTGGGR 643
QY 270 AEYLS---YDSDPC--PGQFTCRGCIKELRCDGMADCTHSDDELNCSDAHQFTCK 325
DB 644 ANFTTGYHIGIRPCEDNPFQGENGECVLLVNLCDGFSKXGSDAHNCV----- 693
QY 326 NKFCKPLFVWCDSDVNDGSDNDEGSCSPAQTFRC--NGKCLSKSQOQCNKDC--G 379
```

Db 694 RF-----LNGTANSG-----LVQFRIGSIWHTNACAEENMTTQSDVQCQLG 735

Qy 380 DGSDBACFP-----KNNVVTCTKHTYRCLNGLCLSKGNBECQDEK-----CSD 423

Db 736 LGTGNSSMPFPSSGGGPFVKLNTAP-----NGSLIITRASBEC--FEDSLILLCNH 784

Qy 424 GSDENDCCGGRSFTFRQA--RVVGGTDADBEEMPMQVSLHLAGGHIICGASLISPNMLYS 481

Db 785 KS-----CGKKQVAQAEVSPKIVGGNDSEGAEMPMVVALYVNGQ-LICGASLVSBDMLVS 837

Qy 482 AAHCYIDRGRFSDPTQMTAFGLIHDSQSRASGVQERRKRIISHPFDFPFEDYDIA 541

Db 838 AAHCYVG----RIIEPSMKKAILGLHMTSNLTSPPQIVTRLDELYINPHYRRKSDIA 893

Qy 542 LLELEKPAEYSSMWRPCLPDASHVFPAGKAIWTTGHTQYGGTALLIKGEIRVINO 601

Db 894 MMHLLEFKNNVYDIQPICLPBNQVFPGRICSLAGMGKVIYQSPADIIQADAVPLISN 953

Qy 602 TTCENLPLQ-QITFRMVCVGLSGGVDSQCDSDSGFPLSSVADGRIPQAGVIVSGDCQA 660

Db 954 EKCCQOMPEYVITENMMCAGYEEGIDSCQDSDSGPLMCLE--NNRMILAGVTSFGYCAL 1012

Qy 661 RNKPGVYTRPLPFRDMK 678

Db 1013 PNRGCVARVKEFTFEMIQ 1030

RESULT 3

A43090

enteropeptidase (EC 3.4.21.9) precursor [validated] - bovine

N:Alternate names: enterokinase

C:Species: Bos primigenius taurus (cattle)

C:Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text change 09-Jul-2004

C:Accession: A43090; A48874; A61436

R:Kittameto, Y.; Yuan, X.; Wu, Q.; McCourt, D.W.; Sadler, J.E.

Proc. Natl. Acad. Sci. U.S.A. 91, 7588-7592, 1994

A:Title: Enterokinase, the initiator of intestinal digestion, is a mosaic protease composed of two distinct domains

A:Reference number: A43090; MUID:94329561; PMID:8052624

A:Accession: A43090

A:Status: nucleic acid sequence not shown; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1035 <KIT>

A:Cross-references: UNIPROT:P98072; GB:U09859; NID:g746410; PID:AA640026.1; PID:g7464111

A:Experimental source: small intestine

R:LaValliere, E.R.; Rehmsmiller, A.; Racie, L.A.; DiBlasio, E.A.; Perent, C.; Grant, K.L.; J. Biol. Chem. 268, 23311-23317, 1993

A:Title: Cloning and functional expression of a cDNA encoding the catalytic subunit of bovine enterokinase

A:Reference number: A48874; MUID:94043122; PMID:8226855

A:Accession: A48874

A:Molecule type: mRNA

A:Residues: 801-1035 <LAV>

A:Cross-references: GB:IJ9663; NID:g416131; PID:AA16035.1; PID:g416132

A:Note: parts of this sequence, including the amino end of the mature protein, were confirmed by sequencing of complementary DNA

R:Light, A.; Janaka, H. J. Protein Chem. 10, 475-480, 1991

A:Title: The amino-terminal sequence of the catalytic subunit of bovine enterokinase. Amino acid sequence and N-terminal amino acid sequence

A:Reference number: A61436; MUID:92189715; PMID:1799406

A:Accession: A61436

A:Molecule type: protein

A:Residues: 801-807, Y, 809-827 <LIG>

C:Comment: The mechanism of association with the membrane of the intestinal brush border membrane attachment using a signal-anchor sequence.

C:Comment: Conversion from membrane-bound to soluble forms may involve further processing

C:Keywords: glycoprotein; hydrolase; intestine; serine proteinase; transmembrane protein; trypsin

F:52-117/Product: enteropeptidase mini chain #status predicted <TM>

F:118-800/Product: enteropeptidase heavy chain #status predicted <HC>

1:fold linked

C:Function: 1. cleaves propeptide from trypsinogen to produce active trypsin

A:Pathway: intestinal digestive hydrolase cascade

C:Superfamily: enteropeptidase; C1/C1s repeat homology; LDL receptor ligand-binding repeat

C:Keywords: glycoprotein; hydrolase; intestine; serine proteinase; transmembrane protein; trypsin

F:52-117/Product: enteropeptidase mini chain #status predicted <TM>

F:118-800/Product: enteropeptidase heavy chain #status predicted <HC>

[illegible]

	A>Title:	Enterokinase, the initiator of intestinal digestion, is a mosaic protease compo
	A Reference number:	A43090; MUID:94329561; PMID:8052624
	A Accession:	B43090
	A Status:	nucleic acid sequence not shown
	A Molecule type:	mRNA
	A Residues:	749-1019 <K12>
	A Cross-references:	GB:U09860
	C Comment:	The mechanism of association with the membrane of the intestinal brush border outlined below) or with amino-terminal myristoylation of the heavy chain.
	C Genetics:	
	A Gene:	GDB:PRSS7
	A Cross-references:	GDB:384083; OMIM:226200
	A Map position:	21q21-21q21
	C Complex:	Mature enteropeptidase is variously reported to contain two (heavy and light) chains by a disulfide bond. Possibly, conversion from membrane-bound to soluble forms involv-
	duga.	
	C Function:	
	A Description:	cleaves activation peptide from trypsinogen to produce active trypsin
	A Pathway:	Intestinal digestive hydrolyase cascade
	C Superfamily:	enteropeptidase; C1r/C1s repeat homology; LDL receptor ligand-binding rep
	C Keywords:	glycoprotein; hydrolase; serine proteinase; transmembrane protein; zymogen
	F 1-784/Product:	enteropeptidase heavy chain #status predicted <CH>
	F 2-38/Domain:	transmembrane #status predicted <TM>
	F 168-221/Domain:	LDL receptor ligand-binding repeat homology <LDL1>
	F 342-504/Domain:	MAM homology <MAM>
	F 526-631/Domain:	C1r/C1s repeat homology <C1R>
	F 643-677/Domain:	LDL receptor ligand-binding repeat homology <LDL2>
	F 678-783/Domain:	scavenger receptor cysteine-rich domain homology #status atypical <SR>
	F 785-1019/Product:	enteropeptidase light chain #status predicted <CH>
	F 785-1014/Domain:	trypsin homology <TRY>
	F 116,147,179,328,335,388,440,470,503,534,630,682,706,725,848,887,909,949/Binding site:	
	F 772-886,810-826,910-977,941-956,967-995/Disulfide bonds:	#status predicted
	F 825,876,971/Active site:	His, Asp, Ser #status predicted
	Query Match	18.0%; Score 683.5; DB 1; Length 1019;
	Best Local Similarity	32.5%; Pred. No. 2,1e-37;
	Matches 180;	Conservative 80; Mismatches 202; Indels 91; Gaps 20;
Oy	161 QLPMSSCGG--RLRKAGTNSPYRGHPNIDCTWNIEVNNQHVKEFKFYLLP	218
Db	519 ELP--TDGSGPEELMEPPTTFSSSTNFPSYNLAFCWILNAOKKNIQLFHEFDLEN	575
Oy	219 RRAOGTCPKDVEI-NGEK-----YCGRSQFVTWSNSNKITVPFHSQSITDPGL	269
Db	576 -----INDVEIRDEADSLAVYTGGCPKVDFTSNKRIVTLITNDVLARGCFK	628
Oy	270 AEYLS---YDSSDPGP-QFTCRGTGRCLRKLRCGDMAICTDSDELNCSGDAHQF---	322
Db	629 ANFTTYGHLGIPEPCAKDHFOCKNGECVPLNLCDGHLCEDSGDEADV-----RFPNG	683
Oy	323 TCRN----KFKCKPLFWCVDSVNDCGNSDEGCSCPATGFSCSNGKCLSKQGCGDDC	378
Db	684 TTNNNGLVRFRIQSII--HTACAEN-----WTQISNDVC-----QLL	719
Oy	379 GDGSDEASCPKVVNVCTKHYTRCLNGCLSKGRPECDG-----KEDCSGDS-----D	426
Db	720 GLGSGNSKPIEST-----DGGPFKLTAPARDGHILLPQQCLODSLIRLOCN	768
Oy	427 EKDCDCGLRSTRQARVVGCTDADBGEMPMOVSLHALGQGHICGASLISPMTVSAHCV	486
Db	769 HKSQKKLAADITPKIYGGNSNAKEGMPPMVVGY-YGGRLLCASLISSDMTLSAHCV	827
Oy	487 IDDRGFRYSDDTWAFGLHDOSORSAPOGBEURLKRTISHRPFNTPPYDIALLELE	546
Db	828 YG---RNLBSKWTAILGLMKSNILTSPOVPRPLIDEIVINPNRRKNDNPANMLE	883
Oy	547 KPAEVSSWARDICLEDASHVEPAGKAIIWTSMGHTOYGEGTALLIKGEIRVINQTCCN	606
Db	884 FKANYTVDIQICLPEENQVFPPPGRNCSIAGMGTVVVOGTANTILEADVFLSNERCQQ	943
Oy	607 LLPO-QITPRMCYGLSGSYDSCQSDSGGSLSSVEADGRIFQAGVVSWGEGCAQRKKPG	665
Db	944 QMPENITENNIIKAGYEEGGIDSQGDSDGGLMQCE--NNRFFLAGVYSFGKCALPBRPG	1002

QY	666	VTYRLFLPFDWIK 678	:: :
Db	1003	VYARVSRFTWIIQ 1015	
RESULT 5			
	JE0315		
	low-density lipoprotein receptor-related protein - mouse		
	C:Species: Mus musculus (house mouse)		
	C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 09-Jul-2004		
	C:Accession: JE0315		
	J:Tomita, Y.; Kim, D.; Magoori, K.; Fujino, T.; Yamamoto, T.T.		
	J:Biochem. 124, 784-789, 1998		
	A:Title: A novel low-density lipoprotein receptor-related protein with type II membrane		
	A:Reference number: JE0315; MUID:98429596; PMID:9756624		
	A:Accession: JE0315		
	A:Status: preliminary		
	A:Molecule type: mRNA		
	A:Residues: 1-1113 <TOM>		
	A:Cross-references: UNIPROT:Q9Z319; DDBJ:AB013874; NID:q3869144; PIDD:BA434371.1; PID:g31		
	C:Superfamily: mouse low-density lipoprotein receptor-related protein; LDL receptor ligand		
	F:337-372/Domain: LDL receptor ligand-binding repeat homology <LDL1>		
	F:374-408/Domain: LDL receptor ligand-binding repeat homology <LDL2>		
	F:410-445/Domain: LDL receptor ligand-binding repeat homology <LDL3>		
	F:447-482/Domain: LDL receptor ligand-binding repeat homology <LDL4>		
	F:648-682/Domain: LDL receptor ligand-binding repeat homology <LDL5>		
	F:684-720/Domain: LDL receptor ligand-binding repeat homology <LDL6>		
	F:723-757/Domain: LDL receptor ligand-binding repeat homology <LDL7>		
	F:863-1097/Domain: trypsin homology <TRI>		
Query Match	17.5%;	Score 663.5;	DB 2; Length 1113;
Best Local Similarity	32.7%;	Pred. No. 4.7e-36;	
Matches 158;	Conservative 67;	Mismatches 187;	Indels 71; Gaps 16;
QY	256	RPHSDQSTTDGFLAEIYSYDSSDPGQFTGTGCIKRLACDGMADCTHSDLNC	315
Db	627	QPEPSSINQITCLP---NEDVECSPSHFKRSGCVLSRRCDGADDDSDSEBNC	683
QY	316	CDAGHQPFCK-NKFKPLFWGCSVNDGDNSEQCS-CPAQTFRCSNGKLSKQCN	373
Db	684	CKERALWCSPPFKQCLKTKTLICDGFPCDSMDKNCSCFQDNELCANHECVPRDLMD	743
QY	374	GKDDCGDSDSEASCPEVN-----VVTCTKHTYRCINQ-----LC-----	407
Db	744	GMVDCSDSDSEMGCVTLISKNGSSILLTVHKSAKEHVCADGWRFTLSQACQMGLEP	803
QY	408	-LSKNGPECDGKE-----DCSDGSDDEKD-----C---DCGLASF	438
Db	804	SVTKILPQEGQGMRLYPNMENLNGSTLQELLLVYHSCPSRSEISLLCKKODGRPPAA	863
QY	439	R-QARVVGCTADDEGEMPMQVSLHLLGQHICGASLISPMVLASAHCYIDDRFRYS	497
Db	864	PMNKRIILGGRSRPRKPMWQCSLQSEPSGHICCVLIAKKVVLTVAHCF-EGR---EDA	918
QY	498	TQWTAFLGHDQSORAPGVQERRLKRILSHPFNFDFEDYIALLELEKPAEYSSNVR	557
Db	919	DVMKVVFGINNLDHDSG-FMQTRPVKTIILHPRYSRAVVDYDISVELSDINETS	977
QY	558	ICLPDASIVFPAGKAIWVTGMGHTQYCGTALILQGEIRVINQTCENILPQO-TTPRM	616
Db	978	VCLSPSEBYLBPDIYCYITGMGH--MGNMPPKLDGEVRIILPLEOCQASFDMKTTINRM	1035
QY	617	MCVGLSGVNSCGDSSGGLPSVLEADGRIFAQGVVSWGDAQOR-NKPGVYRLPLFRD	675
Db	1036	ICAGTESITVDSCHWDSSGGLVCEPFGGQWTLFGLTSMGVSVCPSKVLGPGVYISVSTFVG	1095
QY	676	WIK 678	
Db	1096	WIE 1098	
RESULT 6			

T30337
polyprotein - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C:Accession: T30337
R:Yang, J.C.; Lindsey, L.L.; Hedrick, J.L.
submitted to the EMBL Data Library, March 1998
A:Description: cDNA cloning of ovocytinase, a chymotrypsin-like protease released from XE
A:Reference number: Z20829
A:Accession: T30337
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1524 <YAN>
A:Cross-references: UNIPROT:O91674; EMBL:U81290; NID:92981640; PID:92981641; PIDN:AA0247
C:Superfamily: tryosin related polypeptide; trypsin homology

Query Match 15.2%; Score 576.5; DB 2; Length 1524;
Best Local Similarity 25.9%; Pred. No. 3.5e-30;
Matches 165; Conservative 91; Mismatches 206; Indels 175; Gaps 22;

QY 66 YPARACQWALRGADSVLSLFRFDL---ASCDERSDLVTVNTLS-EMERHALVOL 121
DB 333 YSINSVCCWMLAVOKATIEIRFQLODIEDHATC-----TFDYLSFTVNEKMKIRK 383
QY 122 CGTVPSPYNTLTFHSSQVNLITLITNTERRRHGFEPATFPOLP--RMSGCG-GRLEKAQGT 178
DB 384 CGSTTPSLIY---RSNKVYTFPSDGFRTGREGFIQFLAIPTRAKASCGSAKILIKKGM 440
QY 179 FNSPYPHYPPNIDCTNIEVNNQHVAFKPFYLLPFRACCTCPDYEI-NG--- 234
DB 441 IYSPPYPPYPLRLKTCSSWIEAPENHIVKLFEDENV---EYGHGCIYDAVEYDGAEE 496
QY 235 ----EKYGGERSQPVNTSNKITYRHFSDSYDTGTGLARYLSVDSDPGCGPTCTG 290
DB 497 KQIARLCGYTLPLPISSPENTMIRFTDGENSYGPKFVKF----- 538
QY 291 RCIRKELRCQWADCTDSDDELNCSGADGHOFTCKNK-FCKPLFWVCDVNDGSDNDQ 349
DB 539 ----SFKREKQFSLPV----- 551
QY 350 GCSGPAQTFRCNSGKLSQCGCKGDDCGSDSEASCPKVVNVYTCRHYRCINGCLSL 409
DB 552 ----DD-----TFPISML----- 560
QY 410 KGNPECCKEKDSCDSEKDDCCGARSFTQ---ARVVGCTADBDGEMPMQVSLHALQCG 466
DB 561 --HPRALAD-----VCGMAPMTPKMWLPRIYVGESEASPSNPWVOIIFLRTF 607
QY 467 HICGASLSPMMLVSAHCYIDNRGFRYSPTQMTAFGLHDQSORASAPGVQERLKRIL 526
DB 608 H-CEGAISFQWILTAHC-----IRAAEBSYMTVINGDHKMLNEST-EQIRNIKITR 659
QY 527 SHPFNDFTFDYDIALLEKPAEYSSMWRPICLPDASHVPRAGKAIWVTMGHTQYXGT 586
DB 660 IHDVNSSETYNDIALALYLEBPRLDLDNFVRVCLPEPEPREVLTPASVCVVTCMGNTAEDQ 719
QY 587 GALLQGEIIVNQTCE-NLLPQGITPRMCGVFLSG-GVDSQGGSGGFLSSVAVDG 644
DB 720 PALGLOQLPLRLISICNTSYSGELTDHMLCAGFPSSKSKKACQGGSDGSLVCONEKE 779
QY 645 RIFQAGVVMGDCGQNRKPGVYTRLPPLFRDMIKENT 681
DB 780 QFSYGLVSWGEGCGRVSKPGVYTKVRLFPFWI-ONT 815

RESULT 7
154763
Ra-reactive factor (BC 3.4.21.-) 1 precursor - human
N:Alternate names: mannose binding protein-associated serine proteinase 1 (MASP-1)
C:Species: Homo sapiens (man)
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 09-Jul-2004
C:Accession: I54763; JN0883
R:Sato, T.; Endo, Y.; Matsushita, M.; Fujita, T.

Int. Immunol. 6, 665-669, 1994
A:Title: Molecular characterization of a novel serine protease involved in activation of
A:Reference number: I54763; MUID:94289349; PMID:8018603
A:Accession: I54763
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-699 <SAT>
A:Cross-references: UNIPROT:P48740; GB:D28593; NID:9790963; PIDN:BA05928.1; PID:9471128
R:Takada, P.; Takayama, Y.; Hatanoue, H.; Kanakami, M.
Biochem. Biophys. Res. Commun. 196, 1003-1009, 1993
A:Title: A new member of the C1s family of complement proteins found in a bactericidal f
A:Reference number: JN0883; MUID:94059062; PMID:8240317
A:Accession: JN0883
A:Molecule type: mRNA
A:Residues: 1-234, 'E', 236-284, 'G', 286-498, 'K', 500-542, 'K', 544-642, 'S', 644-699 <TAK>
A:Cross-references: DDBJ:D17525; NID:94339712; PIDN:BA04477.1; PID:94339713
A:Experimental source: liver
C:Comment: This is a serum bactericidal factor that activates complement C4 and C2 compo
C:Genetics:
A:Gene: GDB:MAGP1; GDB:CRAP1; CRAP1; PRSS5; MASP
A:Cross-references: GDB:361104; GDB:330954; OMIM:600521
A:Map position: 3q27-3q28
C:Superfamily: complement-activating serine proteases C1r/C1s/MASP, C1r/C1s repeat homol
C:Keywords: beta-hydroxyasparagine; complement pathway; duplication; glycoprotein; hydro
F:1-17/DNA: signal sequence #status predicted <SIG>
F:18-448,449-699/Product: Ra-reactive factor #status predicted <MAT>
F:19-135/DNA: C1r/C1s repeat homology <C1R1>
F:143-181/DNA: EGF homology <EGF>
F:185-294/DNA: C1r/C1s repeat homology <C1R2>
F:301-362/DNA: complement factor H repeat homology <FH1>
F:367-432/DNA: complement factor H repeat homology <FH2>
F:449-691/DNA: trypsin homology <TRY>
F:49-178,407/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:73-91,143-157,153-166,168-181,185-212,242-260,301-349,329-362,367-414,397-432,436-572,
F:159/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted
F:448-449/Cleavage site: Arg-Ile (autolytic) #status predicted
F:490,552,646/Active site: His, Asp, Ser #status predicted

Query Match 14.9%; Score 565; DB 1; Length 699;
Best Local Similarity 24.9%; Pred. No. 9.1e-30;
Matches 190; Conservative 112; Mismatches 272; Indels 188; Gaps 31;

QY 31 SKTVQRTQDNSCSGLHARGVELMRFTTGPDPSPYPAHARQWALRGADSVLSLFRS 90
DB 15 SKASAHVVELNMMG-----QIOSPGVPDS-YSDSEVTNITVPPGRIKLYFMH 64
QY 91 FDLASCDERSDLVTVNTLS-EMERHALVOLCG-----TYPSPY-NLTFH 134
DB 65 FNLSSYLCEXDYVKV-----ETEDVLAFTCGHETDTQTPQGEVVLSPGSPMSITFR 119
QY 135 GSONVLLITLITNTERRRHGFEPATFPOLP----- 163
DB 120 SD-----FNSBEFTFPFDHYMAVDVDECKERDEBELSCDHYCHNYIGYYCSR 170
QY 164 -----RMSGCGRLRKAQGTENSPPYPCGYPPNIDCTNIEVNNQHVAKVAFKF 212
DB 171 GYLITNTRCRVCSNLTQRTGVTSDPFPPIYKSECLYTLIEBGFVNNLOPED 230
QY 213 FYLLPFRACGTCCKDYVEIN-GEK---YCGERSQPVNTSNKITYRHFSDQSYDTG 267
DB 231 IFDIQDHPV-PCPYDYIKIKVGPVGLPFCGEKAPRISGTQSHSVILFPHSDNSAENRG 289
QY 268 FLAEYLSYDSDPDP-----GQFTKRTGRCIRKELRCQWADCTDSDDELNCSGAGH 320
DB 290 WRLSYRA--AGNECFELQPPVHGKIEPSQAKYFEK-----DQVLVSCDTGY 333
QY 321 QF-----TCKNFKFCPLFM-----VCDVNDGSDNDSEGGSCSPAQ-----TFRCS 361
DB 334 KVLKQNVEMTFOELCAKQGTWNSKIPCTKIVD-----CRAGELHGLITSTR 383
QY 362 NGKLSQCGCKGDDCGSDSEASCPKVVNVYTCRHYRC-----LNGCLSKGNPEC 415
DB 384 NNLITTYSEI---KYSCQF-----PYVGLNNTGTIYCSAQGVMMNKV-LGRSLPTC 432

QY 416 DGEDCDSDGDEKDCDGLSFTRO--ARVVGTDADGEMPMQVSLHALGQGHICGASL 473
D 433 LPV-----CGLPFRSKRLMARIFNGRPAGKGTTPWITAMLSHNGCFCCGSL 479
QY 474 ISPNWLVAHACY--ID--DRGRYSD--PTWMTPLFLGHDSQGSAPGVERRLRKI 525
D 480 LGSWMITVTAHCHLQSDPDGPTLRDSDLSPSDFKILLGKH--WRLKSDENEQH/GVYKH 538
QY 526 ISHFFDFDFDYDIALLEKPAEYSSMWPRICLPASHVFPAGKAIWYMGCHTOYGG 585
D 539 TLHQYVPTNFENDVALVELLESFVLAFWPILCPGPP--QEGANVYISGNG-KQPLQ 595
QY 586 TGAIILOKGEIRVINQTTCENL--LPQQLTPRMVCVFLSGVDSQCGDGGPLSSVEA 642
D 596 RPFETLMEIPIYDHTSQKAVAPLKKKVTBMDICAGEKKGKADACAGDSGGPMVTLNR 655
QY 643 D-GRIFQAGVSWDGCAGRNKPGVYTRLPFRWIKENTGV 683
D 656 ERGQWYLVTGVSWGDGCKDKRGVSVYIHNNKDMIRVTGV 697

RESULT 8

KOMPL
plasma kallikrein (EC 3.4.21.34) precursor - mouse
C/Species: Mus musculus (house mouse)
C/Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 09-Jul-2004
C/Accession: A36557
R/Seidah, N.G.; Sawyer, N.; Hamelin, J.; Mion, P.; Beaudien, G.; Brachpapa, L.; Rochemont
DNA Cell Biol. 9, 737-748, 1990
A/Title: Mouse plasma kallikrein: cDNA structure, enzyme characterization, and comparisc
A/Reference number: A36557; MUID:91090844; PMID:2264928
A/Accession: A36557
A/Molecule type: mRNA
A/Residues: 1-638 <SEI>
A/Cross-references: UNIPROT:P26262; GB:M58588; NID:g200358; PID:AAA6393.1; PID:g200358
A/Note: part of this sequence, including the amino ends of both the heavy and light chain
C/Comment: This protein, synthesized in the liver, circulates as a noncovalent complex w
C/Comment: The zymogen is activated by factor XIIa, which cleaves the molecule into a 11
are linked by one or more disulfide bonds.
C/Superfamily: coagulation factor XI: trypsin homology
C/Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; infla
F/1-19/Domin: signal sequence #status predicted <SIG>
F/20-300/Product: plasma kallikrein heavy chain #status experimental <HCH>
F/110-199/Domin: apple repeat <AP1>
F/200-289/Domin: apple repeat <AP2>
F/291-380/Domin: apple repeat <AP3>
F/391-638/Product: plasma kallikrein light chain #status experimental <LCH>
F/391-621/Domin: trypsin homology <TRY>
F/21-104,47-77,51-57,111-194,137-166,141-147,201-284,227-256,231-237,292-375,318-347,322
F/127,215,308,396,494/Binding site: carbohydrate (asn) (covalent) #status predicted
F/434,483,578/Active site: His, Asp, Ser #status predicted

Query Match 14.4%; Score 545.5; DB 1; Length 638;
Best Local Similarity 27.5%; Pred. No. 1.66-28;
Matches 186; Conservative 87; Mismatches 228; Indels 175; Gaps 32;

QY 99 RGSGLVAVVNTLSMEPHALVQLGCTYPPSYNLFFHSQVLTLLTNTERRHRGP-E 156
D 32 RGGDLAAIYTP---DAQYCKMCTFHRCLLSF-----LAVTPEKTKRPGCFWKE 81
QY 157 ATPQLPRMSSCG--GLRKAQGTFSPPYGPYPP-----NIDCTWNIE-----V 200
D 82 SITGLPRIRHTGASISGHSLKQCHQISACHRDYKGLDRGSMFNISKTDNIECCQKLC 141
QY 201 PNNQHVAVRRKF---FYLLPPRPAC-----GT----- 224
D 142 TNNFHCOF-FYVATSAFYRPRYRKCKLKHASAGSTPISKADNLVSGFSIKSCALSEIG 200
QY 225 CPKO-----YVEINSGKYGERS---QFVYTSNKNITVRFHSDQSYTDG-----FLAE 271
D 201 CPMDIFQHSAPADNLVSCVITTPDAFVCRKTICTFHPNCLFTFTYTNWETSQNRVCFLKT 260

QY 272 YLSYDSDPCPGQ-----FTCRTRG---CIRKEIRCDGMADCTDHSDELNCSCDAGH 320
D 261 SKSGRSPPTIPGMALISGVSLTLCRKRTRPERCHSKI-----YGVDFEGBELVNTFVQGA 315
QY 321 ---QFTCKNFKCPLFLWVCDSV--NDCGNSDEGGCCSPAOTFFCSNGKLSKQCCNGKD 376
D 316 DVCQETC-TTITICQFFIYSLLPQDC---KEGCKXC--SLRLST----- 353
QY 377 DCGDGSDEACPKAVNVYTCFKHTYRCANG-----LCSKNGPECKGKEDSDGDEKDC 430
D 354 ---DGSF-----TRITYGMQSSGYSRLRCLKLVSDCTTKIN----- 388
QY 431 DCGLRSEFTROARVVGTDADGEMPMQVSLHA--LQGHICGASLISPNWLVAHACYID 488
D 389 -----ARIVGFGTNSLGEPMQVSLQVLTAVSQTHLCGSSIIIGQVLTFAHCF-- 436
QY 489 DRGFRISDPQWTAFLGLHDQSQ--RSAPGVERRLRKIISHFFPNDFTDYDIALLELE 546
D 437 -DGIPIY--PDVWRIVYGGIISLSEITKETP---SSRIKELIHQEVKVBEGNDYDIALIKQ 490
QY 547 KPAEYSSMWPRICLPASHVFPAGKAIWYMGCHTOYGTGALILQKGEIRVINQTTCE 606
D 491 TPLANTFRQKPICLPSKADNTITNTNCWYTGWYTKEGSETONILQKATIPLPVNECQK 550
QY 607 LRPQO-ITPRMVCVFLSGVDSQCGDGGPLSSVADGRIFGAGVSWDGCAGRNKPG 665
D 551 KYRDVYINKQMICAGYKGGTADCKDGGPL-VCKHSGRWQVLVGTISWEGCGRRKQPG 609
QY 666 VYTRLPFRWIKENT 681
D 610 VYTKVSEYMDMILEKT 625

RESULT 9

KOMPL
plasma kallikrein (EC 3.4.21.34) precursor - rat
N/Alternate names: Fletcher factor; kininogenin; serum kallikrein
C/Species: Rattus norvegicus (Norway rat)
C/Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 09-Jul-2004
C/Accession: A39180; A33320; S06851; I53041; S06852
R/Beaudien, G.; Rosincki-Chupin, I.; Mattei, M.G.; Molkay, M.; Chretien, M.; Seidah, N.G.
Biochemistry 30, 1628-1635, 1991
A/Title: Gene structure and chromosomal localization of plasma kallikrein.
A/Reference number: A39180; MUID:91129236; PMID:1993380
A/Accession: A39180
A/Molecule type: DNA
A/Residues: 1-638 <BEA>
A/Cross-references: UNIPROT:P14272; GB:J05315
A/Note: the authors translated the codon GAG for residue 81 as Gln
R/Seidah, N.G.; Ladenheim, R.; Molkay, M.; Hamelin, J.; Lutfalla, G.; Rougeon, F.; Lazure
DNA 8, 563-574, 1989
A/Title: The cDNA structure of rat plasma kallikrein.
A/Reference number: A33320; MUID:90091743; PMID:2598771
A/Accession: A33320
A/Status: not compared with conceptual translation
A/Molecule type: mRNA
A/Residues: 1-638 <SEI>
A/Cross-references: GB:M30282; NID:g205010; PID:AAA41463.1; PID:g205011
A/Note: part of this sequence, including the amino ends of both the heavy and light chain
R/Peguin, J.; Benjannet, S.; Sawyer, N.; Lazure, C.; Chretien, M.; Seidah, N.G.
Biochim. Biophys. Acta 999, 103-110, 1999
A/Title: Rat plasma kallikrein: purification, NH(2)-terminal sequencing and development
A/Reference number: S06851; MUID:90089457; PMID:2597701
A/Accession: S06851
A/Molecule type: protein
A/Residues: 20-45;391-413 <PAQ>
R/Seidah, N.G.; Ladenheim, R.; Molkay, M.; Hamelin, J.; Lutfalla, G.; Rougeon, R.; Lazure
DNA Cell Biol. 8, 563-574, 1989
A/Title: The cDNA structure of rat plasma kallikrein.
A/Reference number: I53041
A/Accession: I53041
A/Status: translated from GB/EMBL/DBJ

A:Molecule type: mRNA
 A:Residues: 1-638 <RSS>
 A:Cross-references: GB:M58590; NID:g206721; PIDN:AAA42069.1; PID:g206722
 C:Comment: This protein, synthesized in the liver, circulates as a noncovalent complex with factor XIIa, which cleaves the molecule into a 118 kDa and a 220 kDa subunit.
 C:Comment: The zymogen is activated by factor XIIa, which cleaves the molecule into a 118 kDa and a 220 kDa subunit.
 C:Genetics:
 A:Gene: PK
 C:Superfamily: coagulation factor XI; trypsin homology
 C:Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; infla
 F:1-19/Domin: signal sequence #status predicted <SIG>
 F:20-390/Product: plasma kallikrein heavy chain #status experimental <MAT1>
 F:20-109/Domin: apple repeat <AP1>
 F:110-199/Domin: apple repeat <AP2>
 F:200-289/Domin: apple repeat <AP3>
 F:291-380/Domin: apple repeat <AP4>
 F:391-638/Product: plasma kallikrein light chain #status experimental <MAT2>
 F:391-638/Product: trypsin homology <TRY>
 F:21-104,47-77,51-57,111-194,137-166,141-147,201-284,227-256,231-237,292-375,318-347,322-356,366,384,403,423,443,463,483,503,523,543,563,583,603,623,638/Binding site: carbohydrate (asn) (covalent) #status predicted
 F:434,483,578/Active site: His, Asp, Ser #status predicted
 F:434,483,578/Active site: His, Asp, Ser #status predicted

Query Match 13.9%; Score 526.5; DB 1; Length 638;
 Best Local Similarity 27.4%; Pred. No. 2.9e-27;
 Matches 184; Conservative 85; Mismatches 232; Indels 171; Gaps 32;

Qy 99 RGSGLVTVYNTLSFMEPHALVOLCGTYPPSYNLTFHSSQNVLLTLTNTERRHPGP-E 156
 Db 32 RGGDLAIV--TDAQHC-QKMCCTPHRCCLFSPF-----LAVSPKTEKRGCEFMKE 81
 Qy 157 ATFFQLRMSGCG--GRKRAQGTSPYRPHYPP-----NIDCTNIE-----V 200
 Db 82 SITGTLRIHRTGAIISGLSKQCGHQSACHQDIYEGIDMKSGMNFNISKDISECQKLC 141
 Qy 201 PNNQHV--VRFFFYLLERBRAC-----GT-----C 225
 Db 142 TNNHCGFFTYATATAFHPRERKSCCLKRSSGTPSTIKYPDNLVSGSLKSCALSETGC 201
 Qy 226 PKD-----YVEINGEKYCGERS--QFVTSNSNKTIVRPHSDQSYDTGFLAEYLSYD 276
 Db 202 PMDIFQHPAFADLVNVSQVTPDAFVCRVCTFPHRCCLFPTFTYBEMTESQRNVCFLKTS 261
 Qy 277 SSD-PCF-----GQTCRTGR--CIKELKRDGADCTDHSDELNCSGDAGH- 320
 Db 262 KSGRPSPIIIGENAVSGYSLFTCKRAREPCHFKI-----YSGVAPEGEELNATFVQAD 316
 Qy 321 --OFTCKNKKFCPLFWVCDVNDGDNDSBOGSCSPAOTFRCSNGKCLSKSQCGCKDCC 378
 Db 317 ACQETCTKTRICQFFYTSILPQDC--KAECCK--SLRLST----- 353
 Qy 379 GDGSDASCPKRVNVVCTKATYRCLNGLCLSKGNECDKDCSDGSDKDC-----DC 432
 Db 354 -DGRP-----TRITY-----BAQG-----SSGYSLRCLCKVVSDDC 383
 Qy 433 GLRFTQARVVGTDADGEMPMQVSLHA--LGGGHCASLSLSPMVLVAACFYIDR 490
 Db 384 TTKI---NARIVGGTNSLGEPMQVSLVQVTLVSONHCGSIIIGRWILTAACF--D 437
 Qy 491 GFRSDPTQWTAFLGHDQSO--RSAPGVGRRLKRIISHPFDFTFDYDIALLEKRP 548
 Db 438 GIPY--PDWRIYGGILNLSEITNKP--FSSIELTIHQIKMSBSGSDYIALIKLQTP 492
 Qy 549 AEVSMVRPCLPDASHVFPAGKAIWVGHTQVGTGALILQKGEIRVINOTTCEML 608
 Db 493 LNTYEPQKPLCLPSADTNTIYTNQWVGTGKERGETQNILOKATILVLPNECCQKY 552
 Qy 609 PQQ-ITPRMVCVGLSGVSDSCGDSGAPLSSVEADGAIPOAGVSWDCCQAKRKPVY 667
 Db 553 RQVYITKMTICAGYKEGIDACKDGSGLP--VCKHSGWQDLVGTISWEGCARKEQPGVY 611
 Qy 668 TRLPFRDWIKE 679

Db 612 TKVAEYIDWILE 623

RESULT 10
 complement C3b/C4b inactivator (BC 3.4.21.-) precursor - African clawed frog
 N:Alternate names: C3b/C4b inactivator factor I
 C:Species: Xenopus laevis (African clawed frog)
 C>Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 09-Jul-2004
 C:Accession: 151601; S15468
 R:Kunath-Muglia, L.M.; Chang, G.H.; Sim, R.B.; Day, A.J.; Ezekowitz, R.A.
 M:Immunol. 30, 1249-1256, 1993
 A:Title: Characterization of Xenopus laevis complement factor I structure--conservation
 A:Reference number: 151601; PMID:94019415; PMID:7692240
 A:Accession: 151601
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-613 <KUN>
 A:Cross-references: UNIPROT:Q03711; EMBL:X59958; NID:g64595; PIDN:CAA42582.1; PID:g64596
 C:Superfamily: human complement factor I; LDL receptor ligand-binding repeat homology; tr
 C:Keywords: hydrolase; serine proteinase
 F:220-254/Domin: LDL receptor ligand-binding repeat homology <LDL1>
 F:257-290/Domin: LDL receptor ligand-binding repeat homology <LDL2>
 F:370-599/Domin: trypsin homology <TRY>

Query Match 13.4%; Score 506; DB 2; Length 613;
 Best Local Similarity 26.2%; Pred. No. 6.2e-26;
 Matches 153; Conservative 81; Mismatches 181; Indels 168; Gaps 22;

Qy 225 CPKD---YVEINGEK---YGERSQFVTSNSNKTIVRPHSDQSYDTGFLAEYLSYD 277
 Db 65 CPKATTEVCTDGRKLSYQGLKS--VECSNPLNSKRYRSSBAPCTETFTLQ-----N 117
 Qy 278 SDP-----CEQPTCRTRGCRKELKRDGADCT----- 306
 Db 118 GEPKGIKIKYKLPTEBOELFLCGKQMSREANVVCROLGSTGADASDQVPSLVTEKP 177
 Qy 307 -DHSDELNC-----SC-----DAGH-OFTCKNKKFCPLFW 334
 Db 178 PEHCTIQTGGLNSLAECALRKLPMDONQVAKYCTTENKDCGFGFTSNGKCI9BL 237
 Qy 335 VCDVNDGDNDSBOG--SCPAOTFRCSNGKCLSKSQCGCKDCCGSDGSDASG----- 387
 Db 238 ACDSKNDGSLSDBLCKSKGNA--GFHCRSDPTCIPEQYRCNGELDCIGSEBESNCTVBOQ 296
 Qy 388 -----PKVNVVCTKATYRCLNGLCLSKGNECD 416
 Db 297 KSEKQEVQKQTSKQEDLVQESKATQVEKAKIYNVDIDARRL-----LMSLPE-- 350
 Qy 417 GKEDCSGSDKDCDGL-----RSFTQARVVGTDADGEMPMQVSLHAHQGHICG 470
 Db 351 -----LSCGVPPQTAALTTRKRYVGGTNAVKNQPPMVALKD--GTAVNCG 396
 Qy 471 ASLISPMVLVAACFYIDR-----HDGSDASCPKRVNVVCTKATYRCLNGLCLSKGNECD 525
 Db 397 GIYIGGCVLTAAC-----VSNQPORLYLMEELDRSLSYDDLSFP-----YKSV 444
 Qy 526 ISHPFNDTFPDYDIALLEB-----KPAEYSMVRPCLPDASHVFPAGKAIWVGW 579
 Db 445 IYHETLYNPNTYENDIALLEVYINNPCKMQADNNMPACVPMSPQFKAGDTCTVSGWG 504
 Qy 580 HTQVGGGALILQKGEIRVINOTTCEMLPQOITPRMVCVGLSGVSDSCGDSGAPLSS 639
 Db 505 R-EKGMRSVRFLKKGHTNLND--NCTVYERFLDKKRCARTYDGSIDACKDGSGLPVC 561
 Qy 640 VEAQGRIFQAGVSWDCCQAKRKPVYTRLPFRDWIKE 682
 Db 562 YDVAKVAVWGIYSNGENCGVPGVYTYTKAVYFEMIAQVG 604

RESULT 11
 PLPG
 plasmin (BC 3.4.21.7) precursor - pig (fragment)

N/Alternate names: plasminogen
N/Contains: miniplasminogen
C/Species: Sus scrofa domestica (domestic pig)
C/Date: 07-Sep-1990 #sequence revision 01-Nov-1996 #text_change 09-Jul-2004
C/Accession: S03733; S03737; A25834
R/Schaller, U.; Marti, T.; Roesele, S.J.; Kaempfer, U.; Rickli, E.E.
Fibrinolysis 1, 91-102, 1987
A/Title: Amino acid sequence of the heavy chain of porcine plasmin. Comparison of the ca
A/Reference number: S03733
A/Accession: S03733
A/Molecule type: protein
A/Residues: 1-560 <SCH>
A/Cross-references: UNIPROT:P06867
R/Brunsholz, R.A.; Lerch, P.G.; Schaller, U.; Rickli, E.E.; Lergier, W.; Manneberg, M.;
Eur. J. Biochem. 114, 465-470, 1981
A/Title: Comparison of the primary structure of the N-terminal CNBR fragments of human,
A/Reference number: S03735; MUID:81212097; PMID:7238497
A/Accession: S03737
A/Molecule type: protein
A/Residues: 1-57 <BRU>
R/Marti, T.; Schaller, U.; Rickli, E.E.
Eur. J. Biochem. 149, 279-285, 1985
A/Title: Determination of the complete amino-acid sequence of porcine miniplasminogen.
A/Reference number: A25834; MUID:85203907; PMID:3846533
A/Accession: A25834
A/Molecule type: protein
A/Residues: 450-790 <MAR>
C/Function:
A/Description: dissolves the fibrin of blood clots; acts as a proteolytic factor in a va
ns the walls of the graafian follicle; also activates the urokinase-type plasminogen act
A/Pathway: fibrinolysis
C/Superfamily: plasmin; kringle homology; plasminogen-related protein precursor homology
C/Keywords: fibrinolysis; glycoprotein; hydrolase; kidney; kringle; plasma; serine prote
F.1-790/Product: plasminogen #status predicted <PRO>
F.1-77/Domin: plasminogen-related protein precursor homology (fragment) <PLPH>
F.1-77/Domin: activation peptide #status predicted <APT>
F.78-560/Product: plasmin chain A #status predicted <ACH>
F.84-162/Domin: kringle homology <KR1>
F.166-243/Domin: kringle homology <KR2>
F.256-333/Domin: kringle homology <KR3>
F.358-435/Domin: kringle homology <KR4>
F.450-790/Product: miniplasminogen #status experimental <MIN>
F.461-540/Domin: kringle homology <KR5>
F.561-790/Product: plasmin chain B #status experimental <BCH>
F.561-783/Domin: trypsin homology <TRY>
F.30-54,34-42,84-162,105-145,133-157,166-243,169-229,187-226,215-238,256-333,277-316,305
bonds: #status predicted
F.602,645,740/Active site: His, Asp, Ser #status predicted

Query Match 13.3%; Score 504; DB 1; Length 790;
Best Local Similarity 25.7%; Pred. No. 1.1e-25;
Matches 190; Conservative 80; Mismatches 228; Indels 232; Gaps 35;

74 WALRGDADSVLSLFRSPDLASCDP-----RGSDLVTYNTLSPMN-----P 115
Db WCYTDTDET-----RFYCDIPECEDECHGSEHYEGK-----ISKTSGLHCQSGWSOSP 195
Qy HALVOLGCTVPSPVNLTFHSSQNVLLTLITNTER-----HGCFEATPQLPRMS-- 167
Db HAHGYLPSKF-PNKGLKKNYCRN-----PDGEPRPFCFTTDNKKMEFCDIRCTTPP 247
Qy 168 -----CGRLRKAQGTN-----SPYYPGHVPPNIDCTWNIENVPNQ 204
Db PTSGPTVQCLKRGDENRGVTVASGHTCGRWSAQSPHKNRTPENPPCK-NLE----- 301
Qy 205 HAVKRFKFPYLLERRACGTPKDYVEINRGKYC-----GERSQGVVTSNKKITVR--- 257
Db 302 -----ENYCRNPDGETAPWPCYTTDSE---VRADYVC 328
Qy 258 ---HSDSYTDTGFLAEYLSYSDSDPCGQFTCRGTCIRK-----LRCD 300
Db 329 KIPSCGSSTTST---EHL-----DAVPPEQTPVAQDCYKNGSGSYNGTSTTTTGKRCQ 380

Qy 301 GWADCTDHSDELNCSADGHOFTCKN-----KFCCKPLFWVCDSVNDGDSDEQCS 352
Db 381 SWMSMTPHR-----HEKTPGNFPMNGLTMNVC-----NPDADSPMCYTT 421
Qy 353 CPAQTFR-CSNGKCLSKSQCCNG-----KDDCGDSDDEASCKR-----VAVVT 394
Db 422 DFRVREWCNKKCSSETEQOVNTPPAIAQVSVEDLSSEDCMFGNGKRGKRAITVAGVP 481
Qy 395 C-----TGHTYRCNLGLTSKGNPEC-----DGKED-----CSDGSDKED 429
Db 482 CQEWAGPEHR--HSITPEPTNRAGLEKRYCNRPDGDNDGPMCYTTNPQKLDYCVIPQ 539
Qy 430 C-----DCGRSFTFRQ---ARVVGTDADGEMPMQVSHALQGHICGASLISPMVLVS 481
Db 540 CVMSFPCGKPKVEPKKCPARVVGCVSIPHSWQSLTAYRYRGHPCGGTLISPEWLT 599
Qy 482 AAHCYIDRGRFRSDPIQMTAFGLHDQSRASRPGVERLKLISHPFNDFPDVIA 541
Db 600 AKQCL-----EKSSPSRYKVIIGAHBE-YHAGVGQELDVSKLFXEP-----SEADIA 647
Qy 542 LLELEKPAEYSSMVRPCLPDASHVFPAGKAIWVGHTQYGGTALILQKEIRVING 601
Db 648 LHKSSPAVITDKYIPACLPTFPVYVADRACIITMGERTK-GTYGAGLKEARLPYIEN 706
Qy 602 TTCE--NLPRQITPRMVCVGLSGVDSQGSQSGGLSVSEADRIPOAGVSWGDCA 659
Db 707 KVCNRREYLGKYSVPMELCAGHLAGGIDSCQDSGGPLVCFERDKYTLQ-GVTSWGLGCA 765
Qy 660 QRRKPGYTTLPPLFRDWIKE 679
Db 766 LPMKPGYVAVSRFVWIEE 785

RESULT 12

B61545
Plasmin (EC 3.4.21.7) precursor - sheep (fragments)
N/Alternate names: plasminogen
N/Contains: miniplasminogen
C/Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
C/Date: 28-Oct-1994 #sequence_revision 01-Nov-1996 #text_change 09-Jul-2004
C/Accession: B61545; S28200
R/Schaller, U.; Rickli, E.E.
Enzyme 40, 63-69, 1988
A/Title: Structural aspects of the plasminogen of various species.
A/Reference number: A61545; MUID:89005015; PMID:3168975
A/Accession: B61545
A/Molecule type: protein
A/Residues: 1-37;38-117 <SCH>
A/Cross-references: UNIPROT:P81286
R/Schaller, U.; Straub, C.; Kaempfer, U.; Rickli, E.E.
Protein Seq. Data Anal. 5, 21-25, 1992
A/Title: Complete amino acid sequence of ovine miniplasminogen.
A/Reference number: S28200; MUID:93149995; PMID:1492092
A/Accession: S28200
A/Molecule type: protein
A/Residues: 118-460 <SCH>
C/Superfamily: plasmin; kringle homology; plasminogen-related protein precursor homology,
C/Keywords: fibrinolysis; glycoprotein; hydrolase; kringle; plasma; serine protease; zy
F.1-37,38-117,118-460/Product: plasminogen (fragments) #status experimental <PRO>
F.1-37/Domin: activation peptide (fragment) #status experimental <APT>
F.38-117,118-230,231-460/Product: plasmin (fragments) #status experimental <MAT>
F.41-118/Domin: kringle homology <KR4>
F.418-460/Product: miniplasminogen #status experimental <MIN>
F.132-211/Domin: kringle homology <KR5>
F.226-460/Domin: plasmin chain B #status experimental <BCH>
F.221-453/Domin: trypsin homology <TRY>
F.212,315,410/Active site: His, Asp, Ser #status predicted

Query Match 13.3%; Score 503; DB 2; Length 460;
Best Local Similarity 32.4%; Pred. No. 7.3e-26;
Matches 154; Conservative 49; Mismatches 202; Indels 70; Gaps 18;

Qy 222 CGTCPMDYVEINRGKYCGERSQFV---TSNSKITVRPH-SDGSYTDGFLAEYLSYD 276


```

QY 244 FVTSNSNKTITVRHSDQSYTDNGF-----LAELYSDSSDPG---PQ 284
Db 179 IKVLSTN-----VESGFSKPKCALSEIGHMMIIFQILASPDVAVLTPDA 224
QY 285 FTRTGTCIRKELRCDDMADCTHSDLNSCDAGHFTCKNKRCKLF-----WCD 337
Db 225 FVCTT-----ICTYHP--NC-----LFFPTYNVKIE 250
QY 338 SV-NDCCGNSDEQG---CSCPAQTF-----RCSNGKLSKSQOQNGKDCGDS 383
Db 251 SGRRVCLTKTSESTPSSSTPQENTISGYSLTLTCKRLPCHSKIRP--GVDRGSELN 308
QY 384 BASCPKNNV--TCTK-----HTYRCLNGLSKGNPECDGKEDSGDSDCCDGLR 435
Db 309 VTFVKGNNVCOETCTKMRICQFFYSL-----PEDCK---EEKCKCFLR 350
QY 436 -----SPTROA-----RVNGGTDADEGEMPOVSLHA 462
Db 351 LSMDSFTRLAYGTQSSGSLRLCTGDSNVCCTKTSTRIVGTSWGBWPMQVSLQV 410
QY 463 --LGQGHICGASLISPMMLVSAACVYIDRGFRYSDPTQWTAFLGLHDSQ--RSAPGVQ 518
Db 411 KLRARLRCGSGSLGHQMWLTANCF---DGLPLQD--VWRIYSLNLSDITKDTFSG 465
QY 519 ERRLKRIISHPFNDFFDYDIALLELEKPAEYSSWVRPICLPDASHVFPAGAKIWTGW 578
Db 466 ---IKEIHHQNVKSEGNHDIALIKQAPLNTYEFQKICLPBKDPTSTIYTNCWTGW 522
QY 579 GHYQYGTGALLQKGRIRVYNTTCENLLPQ--QITPRMVCVGLSGVSCGDSGGL 637
Db 523 GFSKEKEIQLNLIKQKNIPLVTNECCQRYQDYITRQWVCAQKXEGKQACKDSGGL 582
QY 638 SSVADGRIFQAGVSWGDCQAKRNGVYTRLPPLFRDWIKENT 681
Db 583 -VCHNGMWRMLVGTISWGBCARREGQVYTKVAEYMDILEKT 625

```

RESULT 15

```

A29154
complement factor I (EC 3.4.21.45) precursor - human
N;Alternate names: C3b/C4b inactivator
C;Species: Homo sapiens (man)
C;Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 09-Jul-2004
C;Accession: A29154; A28434; S66420
R;Catterall, C.F.; Lyons, A.; Sim, R.B.; Day, A.J.; Harris, T.J.R.
Biochem. J. 242, 849-856, 1987
A;Title: Characterization of the primary amino acid sequence of human complement control
A;Reference number: A29154; MUID:87241401; PMID:2954545
A;Accession: A29154
A;Molecule type: mRNA
A;Residues: 1-583 <CAT>
A;Cross-references: UNIPROT:P05156; GB:Y00318
R;Goldberger, G.; Bruns, G.A.P.; Ritz, M.; Edge, M.D.; Kwiatkowski, D.J.
J. Biol. Chem. 262, 10065-10071, 1987
A;Title: Human complement factor I: analysis of cDNA-derived primary structure and assign
A;Reference number: A28434; MUID:87280021; PMID:2956252
A;Accession: A28434
A;Molecule type: mRNA
A;Residues: 1-557,'F',559-583 <GOL>
A;Cross-references: GB:J02770; NID:g182606; PIND:AAA52455.1; PID:g182607
R;Ullman, C.G.; Harris, P.I.; Smith, K.F.; Sim, R.B.; Emery, V.C.; Perkins, S.J.
FEBS Lett. 371, 199-203, 1995
A;Title: beta-Sheet secondary structure of an LDL receptor domain from complement factor
A;Accession: S66420; MUID:95402210; PMID:7672128
A;Molecule type: protein
A;Residues: 258-269 <ULV>
C;Genetics:
A;Gene: GDB:IF
A;Cross-references: GDB:120077; OMIM:217030
A;Map position: 4q24-q25
C;Superfamily: human complement factor I; LDL receptor ligand-binding repeat homology; h
C;Keywords: glycoprotein; hydrolase; serine proteinase

```

```

F;1-18/Domain: signal sequence #status predicted <SIG>
F;19-335/Product: complement factor I heavy chain #status predicted <CFH>
F;221-255/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F;293-293/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F;340-583/Product: complement factor I light chain #status predicted <CFL>
F;340-569/Domain: trypsin homology <TRY>
F;70,103,177,464,494,536/Binding site: carbohydrate (Asn) (covalent) #status predicted

```

```

Query Match 13.0%; Score 492.5; DB 2; Length 583;
Best Local Similarity 30.0%; Pred. No. 4,6e-25;
Matches 133; Conservative 64; Mismatches 189; Indels 57; Gaps 17;

```

```

QY 264 TDTGFLAEVISTSSD---PCPGQTCRGRIRKELRCDDMADCTD--HSDLNSC 316
Db 165 TORRFLSDLSINSTCLHVCRLGT-SLAETFTKRRRMGQDPADVVCYQKADSPW 223
QY 317 DADHPTCKKKPKPLFWCDSDVDCGDSNDEQGC--SCPAQTRGNGKLSQOQNGK 375
Db 224 D--DFQCVNGKTI SQMKACDGINCGDSDELCKACQKGFHCKSGVCIPOYQNGE 281
QY 376 DDCGDSDEASG-----PKVNVYCTKHTYRCLNGLSKGNPECDGKEDSGSDE 427
Db 282 VDCITBEDVCGAFASVAQEETELLTAQMDARRRIKSLPK----- 324
QY 428 KDCDCGL--RSFTROARVYGTDDADGEMPMQVSLHALGCGHICGASLISPMMLVSAAC 485
Db 325 --LSCCVKRMHRRIRRIYGVGRKQAGDLPQVAILD--ASGLICGIGIYGCCWILTAHC 381
QY 486 YIDRGFRYSDPTQWTAFLG-LHDSQSAFQGERLKRRIISHPFNPTFDYDIALLE 544
Db 382 LRASKTHRYD--IWTTVDMIRPDLK---IVIEYDRILIHENYNACTYQNDIALLE 434
QY 545 LEKPAEYS--SMVR--PICLPDASHVFPAGAKIWTGNGHTQYGTGALLQKGRIRVN 600
Db 435 MKDGNKKXOCDELPRSIACVPWSPYLFQPNDRCTIVSGWGR-EKDNRRVESLQGEVGLI- 492
QY 601 QTTCEMLPQQTIPR--MVCYGLSGVDSGCGSGPLSVENDGRIFQAGVSWKDGCA 659
Db 493 -SNCSKRYGNRFYKEMECAGTYDGSIDACKDSGGLVCMANNTYVWGVSWGECNG 551
QY 660 QRKPKGVYTRLPPLFRDWIKENTG 682
Db 552 KPEFPGVYTKVANYFPMISHVHG 574

```

RESULT 16

```

PIHU
plasma (EC 3.4.21.7) precursor [validated] - human
N;Alternate names: plasminogen precursor [misnomer]
N;Contains: angiotatin; microplasmin; plasminogen
C;Species: Homo sapiens (man)
C;Date: 24-Apr-1984 #sequence_revision 02-Dec-1994 #text_change 09-Jul-2004
C;Accession: A35229; A52242; A266738; I84609; S03735; A00929; A04625; A046
R;Peterson, T.E.; Martzen, M.R.; Ichinose, A.; Davie, E.W.
J. Biol. Chem. 265, 6104-6111, 1990
A;Title: Characterization of the gene for human plasminogen, a key proenzyme in the fibr
A;Reference number: A35229; MUID:90202879; PMID:2318848
A;Accession: A35229
A;Molecule type: DNA
A;Residues: 1810 <BET>
A;Cross-references: UNIPROT:P00747; UNIPROT:Q9UBQ9; UNIPROT:Q9UMI2; GB:J05286; GB:M34276;
R;Experimental source: leukocyte; lung fibroblast
R;Malgaroli, N.; Bruno, L.; Pontoglio, M.; Candiani, G.; Meroni, G.; Ottolenghi, S.; Tar
Biochem. Biophys. Res. Commun. 173, 1013-1018, 1990
A;Title: Definition of the transcription initiation site of human plasminogen gene in liv
A;Reference number: I52242; MUID:91097523; PMID:2268308
A;Accession: I52242
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-16 <MML1>
A;Cross-references: GB:M62890; NID:g190092; PIND:AAA6454.1; PID:g553613
R;Forsgren, M.; Raden, B.; Israelsson, M.; Larsson, K.; Heden, L.O.
FEBS Lett. 213, 254-260, 1987

```

A:Title: Molecular cloning and characterization of a full-length cDNA clone for human p
A:Reference number: A26646; MUID:87162490; PMID:3030813
A:Accession: A26646
A:Molecule type: mRNA
A:Residues: 1-471, 'D', 473-810 <FOR>
A:Cross-references: GB:X05199; NID:g35530; PIDN:CAA28831.1; PID:g35531
A:Experimental source: liver
R:Malinowski, D.P.; Sadler, J.E.; Davie, E.W.
Biochemistry 23, 4243-4250, 1984
A:Title: Characterization of a complementary deoxyribonucleic acid coding for human and
A:Reference number: 145961; MUID:85023311; PMID:6148961
A:Accession: 162738
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 292-471, 'D', 473-810 <MAL2>
A:Cross-references: GB:X02922; NID:g190112; PIDN:AAA60124.1; PID:g387031
A:Accession: 184609
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 367-419 <MAL3>
A:Cross-references: GB:X02921; NID:g190110; PIDN:AAA60123.1; PID:g190111
R:Brunisholm, R.A.; Lerch, P.G.; Schaller, J.; Rickli, E.E.; Lergler, W.; Manneberg, M.;
Eur. J. Biochem. 114, 465-470, 1981
A:Title: Comparison of the primary structure of the N-terminal CNBR fragments of human,
A:Reference number: S03735; MUID:81212097; PMID:7238497
A:Accession: S03735
A:Molecule type: protein
A:Residues: 20-71, 'E', 73-76 <BRU>
R:Schirup-Jensen, L.; Petersen, T.E.; Magnusson, S.
submitted to the Atlas, July 1977
A:Reference number: A00929
A:Accession: A00929
A:Molecule type: protein
A:Residues: 20-71, 'E', 73-85, 87-106, 'D', 108-360, 'E', 362-810 <SOT>
R:Wiman, B.
Eur. J. Biochem. 76, 129-137, 1977
A:Title: Primary structure of the B-chain of human plasmin.
A:Reference number: A04627; MUID:77225245; PMID:142009
A:Accession: A04627
A:Molecule type: protein
A:Residues: 581-810 <W11>
R:Wiman, B.; Wallen, P.
Eur. J. Biochem. 50, 489-494, 1975
A:Title: Structural relationship between "glutamic acid" and "lysine" forms of human pla
A:Reference number: A04625; MUID:75093329; PMID:122932
A:Accession: A04625
A:Molecule type: protein
A:Residues: 20-50, 'Q', 51-71, 'E', 73-85, 87-100 <W12>
R:Wiman, B.; Wallen, P.
Eur. J. Biochem. 58, 539-547, 1975
A:Title: Amino-acid sequence of the cyanogen-bromide fragment from human plasminogen the
A:Reference number: A04626; MUID:76043692; PMID:126863
A:Accession: A04626
A:Molecule type: protein
A:Residues: 483-507, 'E', 509-604 <W13>
R:Robbins, K.C.; Bernabe, P.; Arzadon, L.; Summatta, L.
J. Biol. Chem. 248, 1631-1633, 1973
A:Title: The primary structure of human plasminogen. II. The histidine loop of human pla
A:Reference number: A92125; MUID:73149248; PMID:4694729
A:Contents: annotation; active site
R:Groskopf, W.R.; Summatta, L.; Robbins, K.C.
J. Biol. Chem. 244, 3590-3597, 1969
A:Title: Studies on the active center of human plasmin. Partial amino acid sequence of a
A:Reference number: A92048; MUID:69234739; PMID:4240117
A:Contents: annotation; active site
R:Trexler, M.; Vail, Z.; Patchy, L.
J. Biol. Chem. 257, 7401-7406, 1982
A:Title: Structure of the omega-amino-carboxylic acid-binding sites of human plasminogen.
A:Reference number: A92382; MUID:82213905; PMID:6919539
A:Contents: annotation; omega-amino-carboxylic acid binding sites
R:Vail, Z.; Patchy, L.
J. Biol. Chem. 259, 13690-13694, 1984

A:Title: The fibrin-binding site of human plasminogen. Arginines 32 and 34 are essential
A:Reference number: A52458; PMID:85054794; PMID:6094526
A:Contents: annotation; fibrin binding site; omega-aminocarboxylic acid binding site
J:Gao, Y.; Ji, R.W.; Davidson, D.; Schaller, J.; Marti, D.; Soehnlel, S.; McCance, S.G.;
J: Biol. Chem. 271, 29461-29467, 1996
A:Title: Kringle domains of human angiotensin. Characterization of the anti-proliferative
A:Reference number: A58811; PMID:97067211; PMID:8910613
A:Contents: annotation
R:LiJihen, H.R.; Uguw, F.; Bini, A.; Collen, D.
Biochemistry 37, 4699-4702, 1998
A:Title: Generation of an angiotensin-like fragment from plasminogen by streptolysin-1 (WP
A:Reference number: A58812; PMID:9548733; PMID:9548733
A:Contents: annotation
R:Tulinaky, A.; Mulinak, A.M.
submitted to the Brookhaven Protein Data Bank, July 1991
A:Reference number: A51486; PDB:1ZPK4
A:Contents: annotation; X-ray crystallography, 2.25 angstroms, residues 375-454
R:Mu, T.P.; Tulinaky, A.
submitted to the Brookhaven Protein Data Bank, August 1993
A:Reference number: A51911; PDB:1PKR
A:Contents: annotation; X-ray crystallography, 2.48 angstroms, residues 102-181
R:Padmanabhan, K.; Tulinaky, A.
submitted to the Brookhaven Protein Data Bank, April 1994
A:Reference number: A52408; PDB:1PMK
A:Contents: annotation; X-ray crystallography, 2.25 angstroms, residues 377-454
R:Tulinaky, A.; Mathews, I.I.
submitted to the Brookhaven Protein Data Bank, December 1995
A:Reference number: A55244; PDB:1CEA
A:Contents: annotation; X-ray crystallography, 2.1 angstroms, residues 102-181
R:Tulinaky, A.; Mathews, I.I.
submitted to the Brookhaven Protein Data Bank, December 1995
A:Reference number: A65245; PDB:1CEB
A:Contents: annotation; X-ray crystallography, 2.1 angstroms, residues 102-181
R:Mulinak, A.M.; Tulinaky, A.; Ravichandran, K.G.
Biochemistry 30, 10576-10588, 1991
A:Title: Crystal and molecular structure of human plasminogen kringle 4 refined at 1.9 Å
A:Reference number: A58819; PMID:92031502; PMID:1657148
A:Contents: annotation
R:Mu, T.P.; Padmanabhan, K.; Tulinaky, A.; Mulinak, A.M.
Biochemistry 30, 10589-10594, 1991
A:Title: The refined structure of the epsilon-aminocaproic acid complex of human plasmin
A:Reference number: A58818; PMID:92031503; PMID:1657149
A:Contents: annotation
R:de Vos, A.M.; Ultsch, M.H.; Kelley, R.F.; Padmanabhan, K.; Tulinaky, A.; Westbrook, M.J.
Biochemistry 31, 270-279, 1992
A:Title: Crystal structure of the kringle 2 domain of tissue plasminogen activator at 2.1 Å
A:Reference number: A39483; PMID:92118803; PMID:1310033
A:Contents: annotation; X-ray crystallography, 2.4 angstroms
R:Stec, B.; Teeter, M.M.; Whitlow, M.; Yamano, A.
submitted to the Brookhaven Protein Data Bank, June 1995
A:Reference number: A65980; PDB:1KRN
A:Contents: annotation; X-ray crystallography, 1.67 angstroms, residues 376-454
R:Rejante, M.; Llinas, M.
submitted to the Brookhaven Protein Data Bank, August 1996
A:Reference number: A65803; PDB:1HRJ
A:Contents: annotation; conformation by (1)H-NMR, residues 103-181
R:Rejante, M.; Llinas, M.
submitted to the Brookhaven Protein Data Bank, August 1996
A:Reference number: A65804; PDB:1HPK
A:Contents: annotation; conformation by (1)H-NMR, residues 103-181
R:Rejante, M.R.; Llinas, M.
Biochemistry 221, 927-937, 1994
A:Title: (1)H-NMR assignments and secondary structure of human plasminogen kringle 1.
A:Reference number: A43645; PMID:94237157; PMID:818175
A:Contents: annotation; conformation by (1)H-NMR, residues 96-184
R:Rejante, M.R.; Llinas, M.
Biochemistry 221, 939-949, 1994
A:Title: Solution structure of the epsilon-aminohexanoic acid complex of human plasminogen
A:Reference number: A58817; PMID:94237158; PMID:818176

A:Molecule type: DNA
 A:Residues: 1-461 <ANS>
 A:Cross-references: GB:K02048
 R:Retama, P.H.; Bertina, R.M.; Ploos van Amstel, J.K.; Riemens, A.; Briet, E.
 Blood 72, 1074-1076, 1988
 A:Title: The putative factor IX gene promoter in hemophilia B Leyden.
 A:Reference number: A30511; MUID:88327116; PMID:3416069
 A:Accession: A30511
 A:Molecule type: DNA
 A:Residues: 8-24 <RE>
 A:Cross-references: EMBL:X55008; NID:G11288; PIDN:CA838245.2; PID:G4469253
 R:Koebel, D.D.; Botema, C.D.K.; Buerschede, J.M.; Sommer, S.S.
 Am. J. Hum. Genet. 45, 448-457, 1989
 A:Title: Functionally important regions of the factor IX gene have a low rate of polymor
 A:Reference number: A32989; MUID:88371752; PMID:2773937
 A:Accession: A32989
 A:Status: not compared with conceptual translation
 A:Molecule type: DNA
 A:Residues: 30-92 <KOE>
 R:McGraw, R.A.; Davie, L.M.; Noyes, C.M.; Lundblad, R.L.; Roberts, H.R.; Graham, J.B.; S
 Proc. Natl. Acad. Sci. U.S.A. 82, 2847-2851, 1985
 A:Title: Evidence for a prevalent dimorphism in the activation peptide of human coagul
 A:Reference number: A22673; MUID:85190593; PMID:3857619
 A:Accession: A22673
 A:Molecule type: mRNA
 A:Residues: 1-193, 'T', 195-461 <MCG>
 A:Cross-references: GB:M1309; NID:G180552; PIDN:AA52023.1; PID:G180553
 A:Note: the authors translated the codon ACA for residue 29 as Tyr
 R:Jaye, M.; de la Salle, H.; Schamber, F.; Bailand, A.; Kohli, V.; Finkel, A.; Tolstosh
 Nucleic Acids Res. 11, 2325-2335, 1983
 A:Title: Isolation of a human anti-haemophilic factor IX cDNA clone using a unique 52-b
 A:Reference number: A21337; MUID:83220788; PMID:6687940
 A:Accession: A21337
 A:Molecule type: mRNA
 A:Residues: 1-193, 'T', 195-461 <JAY>
 A:Cross-references: GB:J00137; NID:G182610; PIDN:AA52763.1; PID:G182611
 R:Jagadeeswaran, P.; Lavelle, D.E.; Kaul, R.; Mohandas, T.; Warren, S.T.
 Somat. Cell Mol. Genet. 10, 465-473, 1984
 A:Title: Isolation and characterization of human factor IX cDNA: identification of Tag I
 A:Reference number: A37546; MUID:84300526; PMID:6089357
 A:Accession: A37546
 A:Molecule type: mRNA
 A:Residues: 38-193, 'T', 195-326 <JAG>
 A:Cross-references: GB:M35672
 R:Kurachi, K.; Davie, E.W.
 Proc. Natl. Acad. Sci. U.S.A. 79, 6461-6464, 1982
 A:Title: Isolation and characterization of a cDNA coding for human factor IX.
 A:Reference number: A30623; MUID:83065193; PMID:6959130
 A:Accession: A30623
 A:Molecule type: mRNA
 A:Residues: 1-12, 'S', 14-73, 'P', 75-82, 'K', 84-203, 'P', 205-216, 'G', 218-298, 'A', 299-356, 'A',
 A:Cross-references: GB:J00136; NID:G182608; PIDN:AA598726.1; PID:G182609
 R:Tharakan, J.; Strickland, D.; Burgess, W.; Drchan, W.N.; Clark, D.B.
 Vox Sang. 58, 21-29, 1990
 A:Title: Development of an immunofluorescence process for factor IX purification.
 A:Reference number: A60486; MUID:90194857; PMID:2316207
 A:Accession: A60486
 A:Molecule type: protein
 A:Residues: 47-52, 'XX', 55-60, 'X', 62, 'XX', 65 <THA>
 R:McMullen, B.A.; Fujikawa, K.; Kistle, W.
 Biochem. Biophys. Res. Commun. 115, 8-14, 1983
 A:Title: The occurrence of beta-hydroxyaspartic acid in the vitamin K-dependent blood co
 A:Reference number: A20274; MUID:83308813; PMID:6688526
 A:Accession: A20274
 A:Molecule type: protein
 A:Residues: 105-109, 'X', 111-115 <MCM>
 R:Bailand, A.; Faure, T.; Carvallo, D.; Cordier, P.; Ulrich, P.; Fournet, B.; de la Sall
 Eur. J. Biochem. 172, 565-572, 1988
 A:Title: Characterisation of two differently processed forms of human recombinant factor
 A:Reference number: S02527; MUID:88166735; PMID:3280312
 A:Accession: S02527
 A:Molecule type: protein

A:Residues: 29-63 <BAL>
 A:Note: processed forms expressed in recombinant system
 R:Jallat, S.; Perraud, P.; Dalemans, W.; Bailand, A.; Diesterle, A.; Faure, T.; Meulien, I
 EMBO J. 9, 3295-3301, 1990
 A:Title: Characterization of recombinant human Factor IX expressed in transgenic mice and
 A:Reference number: S12058; MUID:91006024; PMID:2209546
 A:Accession: S12058
 A:Molecule type: mRNA: protein
 A:Residues: 1-68 <JAL>
 A:Note: processed forms expressed in recombinant system
 R:Handford, P.A.; Baron, M.; Mayhew, M.; Willis, A.; Beesley, T.; Brownlee, G.G.; Campbe
 EMBO J. 9, 475-480, 1990
 A:Title: The first EGF-like domain from human factor IX contains a high-affinity calcium
 A:Reference number: S12377; MUID:90151623; PMID:2406129
 A:Accession: S12377
 A:Molecule type: protein
 A:Residues: 92-130 <HAN>
 A:Note: NMR detection of calcium binding by domain expressed in recombinant system
 R:de la Salle, C.; Charanfar, J.L.; Baas, M.J.; Schwartz, A.; Wiesel, M.L.; Grunbaum,
 Thromb. Haemost. 70, 370-377, 1993
 A:Title: A deletion located in the 3' non translated part of the factor IX gene responsi
 A:Reference number: I59612; MUID:94054330; PMID:8236150
 A:Accession: I59612
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 444-461 <RES>
 A:Cross-references: GB:S66752; NID:G439773; PIDN:AA828588.1; PID:G439774
 R:Stoffel, E.S.; Koebel, D.D.; Sarkar, G.; Sommer, S.S.
 Science 239, 491-494, 1988
 A:Title: Genomic amplification with transcript sequencing.
 A:Reference number: I59529; MUID:88127096; PMID:3340835
 A:Accession: I59529
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 290-359 <RE>
 A:Cross-references: GB:M19063; NID:G182622; PIDN:AA52456.1; PID:G182623
 R:Agarwala, K.L.; Kawabata, S.; Takao, T.; Murata, H.; Shimomishi, Y.; Nishimura, H.; Iw
 Biochemistry 33, 5167-5171, 1994
 A:Title: Activation peptide of human factor IX has oligosaccharides O-glycosidically lin
 A:Reference number: A54255; MUID:94227047; PMID:8172892
 A:Accession: A54255
 A:Molecule type: protein
 A:Residues: 'P', 204, 'X', 206-211, 212, 'D', 214, 'X', 216-221, 'D' <AGA>
 A:Note: the residues designated 'X' were determined to be threonine bound to carbohydrate
 R:Di Scipio, R.G.; Kurachi, K.; Davie, E.W.
 J. Clin. Invest. 61, 1528-1538, 1978
 A:Title: Activation of human factor IX (Christmas factor).
 A:Reference number: A18483; MUID:78194509; PMID:659613
 A:Contents: annotation; activation; active site; carbohydrate binding
 R:McGraw, R.A.; Davie, L.M.; Noyes, C.M.; Graham, J.B.; Roberts, H.R.; Stafford, D.W.
 Am. Soc. Hematol. Abstr. 64(Suppl.1), 262a, 1984
 A:Reference number: A37569
 A:Contents: annotation
 A:Note: 194-Thr was also found
 R:Morita, T.; Isaacs, B.S.; Esmom, C.T.; Johnson, A.E.
 J. Biol. Chem. 259, 5698-5704, 1984
 A:Title: Derivatives of blood coagulation factor IX contain a high affinity Ca²⁺-binding
 A:Reference number: A37543; MUID:84185715; PMID:6425256
 A:Contents: annotation; calcium binding
 R:Morita, T.; Isaacs, B.S.; Esmom, C.T.; Johnson, A.E.
 J. Biol. Chem. 260, 2583, 1985
 A:Reference number: A37544
 A:Contents: annotation; calcium binding, correction
 R:Bentley, A.K.; Rees, D.J.G.; Rizza, C.; Brownlee, G.G.
 Cell 45, 343-348, 1986
 A:Title: Defective propeptide processing of blood clotting factor IX caused by mutation c
 A:Reference number: A37545; MUID:86189947; PMID:3009023
 A:Contents: annotation; signal sequence cleavage site
 R:Shenito, K.; Kawabata, S.; Miyata, T.; Takeya, H.; Takamatsu, J.; Ogata, K.; Kamiya,
 J. Biol. Chem. 264, 21257-21265, 1989
 A:Title: Blood clotting factor IX B(M) Nagoya: substitution of arginine 180 by tryptophan
 A:Reference number: A30622; MUID:90078229; PMID:2592373
 A:Contents: annotation; sequence of mutant B(M) Nagoya

A>Note: carboxylation, glycosylation, and cleavage sites
R:Baron, M.; Norman, D.G.; Harey, T.S.; Hanford, P.A.; Mayhew, M.; Tse, A.G.D.; Brownle
submitted to the Brookhaven Protein Data Bank, November 1991
A:Reference number: A51252; PDB:1IXA
A:Contents: annotation; conformation by (1)H-NMR, residues 92-130
A>Note: recombinant form expressed in yeast
C:Comment: Factor IX is activated by factor XIa, which excises the activation peptide pr
C:Comment: The gamma-carboxyglutamic acid residues arise by posttranslational, vitamin K
C:Comment: Calcium binds to the gamma-carboxyglutamic acid (Gla) residues and, with strc
C:Genetics:
A:Gene: GDB:F9
A:Cross-references: GDB:119900; OMIM:306900
A:Map position: Xq27.1-Xq27.2
A:Functions: 30/1; 84/2; 93/1; 131/1; 174/1; 241/3; 280/1
A:Function:
A:Description: catalyzes the proteolytic activation of coagulation factor X in the prese
C:Superfamily: blood coagulation intrinsic pathway
C:Keywords: beta-hydroxyaspartic acid; blood coagulation; calcium binding; carboxyglutan
F:1-28/Domain: signal sequence #status predicted <SIG>
F:29-46/Domain: propeptide #status experimental <PPT>
F:31-91/Domain: Gla domain homology <GLA>
F:47-191/Product: coagulation factor IXa light chain #status experimental <ALC>
F:97-128/Domain: EGF homology <EG1>
F:134-170/Domain: EGF homology <EG2>
F:192-226/Domain: activation peptide #status experimental <ACT>
F:227-461/Product: coagulation factor IXa heavy chain #status experimental <AHC>
F:227-454/Domain: trypsin homology <TRY>
F:53,54,61,63,66,67,72,73,76,78,82,86/Modified site: gamma-carboxyglutamic acid (Glu) #
F:64-69,97-108,102-117,119-128,134-145,141-155,157-170,178-335,252-268,382-396,407-435/D
F:99/Binding site: carbohydrate (Ser) (covalent) #status experimental
F:110/Modified site: erythro-beta-hydroxyaspartic acid (Asp) #status experimental
F:191-192/Cleavage site: Arg-Ala (coagulation factor IXa) #status experimental
F:203,213/Binding site: Asn (covalent) #status experimental
F:205,215/Binding site: carbohydrate (Thr) (covalent) #status experimental
F:226-227/Cleavage site: Arg-Val (coagulation factor IXa) #status experimental

Query Match 12.8%; Score 484.5; DB 1; Length 461;
Best Local Similarity 27.6%; Pred. No. 1.2e-24;
Matches 145; Conservative 76; Mismatches 184; Indels 121; Gaps 19;

Qy 188 YPPRIDCTNMIENVNNOHVKRFFFLLEPRACGTCPPDYVINEKTC-GRSQPVV 246
Db YLSAECTVFLDENANKI-----LNRPKRNSGKLEEFVQGNLERRECKSCFEE 73
Qy 247 TSNSNKITVFPHSQSTYDTCFLAEVLSYD--SSDPCPGQFTKRTG----- 290
Db 74 ARVFEENTER-----TFEFKQIVDGDQCESNPLNGSCSKDDINSYECMCPFEFG 125
Qy 291 -----RCIRKELRCGMADCTDHS-D-ELNCSGDAGHPTCKNKKFPLFWVCDVND 342
Db 126 KNCGLDYVTCNKNRGRCPF--CKNSADNKKVCSCTEGVRLAENKSCSE----- 171
Qy 343 GDNDSDEGSCCPAOTFCSCNGKCLSKSQCKGKDCDGDSDASCPKRVNVTCTKHTYRC 402
Db 172 -----PAVFPFCGRVSVQTSKLTTRA-----EAVFPVDVYNSTE----- 206
Qy 403 LINGCLSKNPECGKSDGSDGDEKDCDGLRSTFRQARVVGSTDDEGEPMQVSLHA 462
Db 207 -----AETIIDNTQ-----STGSFNDPTRVGGEDAKPQGFPMQVVLNG 246
Qy 463 LGGGHCIGASLISPMVLVSAACHYIDRGFRYSPTQWTAFLGLHDOSQSAFVQERRLK 522
Db 247 KQDA-FCGGSIVNEKMTVTAHCV--ETGVKI-----TVVAGEHN-IEBTEHTEQGRNV 296
Qy 523 KRITSHPEPNDF--TFPYDIALLELEKPAEYSSMVRPCLPDASHV-----FPAKGIWV 575
Db 297 IRIIPHNNAALINKYHNDIALLELDEPLVNSVYTPICIADEKTYTIPLKFGSG---YV 353
Qy 576 TGMGHTQYGGTALILKGEIRVINOCTCENTLPOOLTPPRMVCVFLSGGVDSQGGSGG 635
Db 354 SGMGKRVHKGKRSALVLOTLRVPLVDRAITCLARSTFTTYNNMFCAGFHEGGKDSQGGSDG 413

Qy 636 PLSSVEADGRIFQAGVSWGDCGQARNKPGVTYRLPLFMDMIKENT 681
Db 414 P-HVTEVEGTSFLTGIIISWGEBCAMKKGKGIYTKVSRVVMIKEXT 458

RESULT 22
I46260
Plasmin (EC 3.4.21.7) precursor - western European hedgehog
C:Species: Erinaceus europaeus (western European hedgehog)
C:Date: 21-Feb-1997 #sequence_rev150n 21-Feb-1997 #text_change 09-Jul-2004
C:Accession: I46260
R:Lawn, R.M.; Boommark, N.W.; Schwartz, K.; Lindahl, G.E.; Wade, D.P.; Byrne, C.D.; Fong,
J. Biol. Chem. 270, 24004-24009, 1995
A:Title: The recurring evolution of lp(a): Insights from cloning of hedgehog apolipoprote
A:Reference number: I46259; MIM:6025778; PMID:7592597
A:Accession: I46260
A:Status: preliminary; translated from GB/EMBL/DBD
A:Molecule type: mRNA
A:Residues: 1-810 <LAW>
A:Cross-references: UNIPROT:Q29485; EMBL:U03171; NID:G1046360; PID:G1046361
C:Superfamily: plasmin; kringle homology; plasminogen-related protein precursor homology;
C:Keywords: hydrolase; serine protease
F:1-96/Domain: plasminogen-related protein precursor homology <PLPH>
F:103-181/Domain: kringle homology <KR1>
F:185-262/Domain: kringle homology <KR2>
F:275-352/Domain: kringle homology <KR3>
F:379-456/Domain: kringle homology <KR4>
F:482-561/Domain: kringle homology <KR5>
F:582-803/Domain: trypsin homology <TRY>

Query Match 12.7%; Score 483; DB 2; Length 810;
Best Local Similarity 26.7%; Pred. No. 2.7e-24;
Matches 170; Conservative 78; Mismatches 205; Indels 184; Gaps 31;

Qy 192 IDC-TNNIRVFN-----NGHYKVRKFFYLLEPRACGTCK-----DVEI----- 232
Db 204 LECQPMDSQIPHPHGFIPSKFPKSNLKNMNCRNPDGSPRCWCTFMDNKRMEYCDIPRCT 263
Qy 233 -----NGEKYCGERSQFVVTNSNKTIVFPHSQS-----YTDG 267
Db 264 TTPPPSGPTVQCLMNGEHNQGN-----VAVVYSGLTQRMGDSPHRDRTPENYPCKN 318
Qy 268 FLAEVLSYDSDPCPGQFTCTGRCIRKEL---RCDGMADCTDHSDE----- 311
Db 319 IDENYCRNPDGEPAPMCF--TNSVSRWECKIPDCVSSASEYHSDAPVTPPEQTPVV 376
Qy 312 LNSCDAHQF-----TCNKRCKPLFWV-----CD-SVNDGNSDE 348
Db 377 QECYQNGQYRTGTSSTTTGKKCP--WTSNRPRHSKTPENYPPADLTMTNCRNPDGD 434
Qy 349 QGSCS-----PAQTFR-CSNKGCL-----SKSQOC--NGKDDCGDG 381
Db 435 KRPWCYTTTPSVAMERCNLKKCGSTEMSATNSSPVQVSASBSBEDCIIDNKGVRGK 494
Qy 382 SDEAS-----CPRVNVTCTKHTYRC-----LINGCLSKNPE--- 414
Db 495 ATTGAGTTCQAMAAPRHRHSITPPTNPRADIQENY-CRNPDGDANGPCTTNRKLP 553
Qy 415 --CDGKEDSGSDGDEKDCDGLRSTFRQARV-----YCGTDADGEPMQVSLHALG 464
Db 554 DYCDIPIHCVPSS--ADCG-----KRPVEPKCPRGVGCVAPHPSPMVSILRRFG 603
Qy 465 QGHCIGASLISPMVLVSAACHYIDRGFRYSPTQWTAFLGLHDOSQSAFVQERRLK 524
Db 604 Q-HFCGTLISPMVVTAAHCL-----EKFSNAIKYVVLGAH-QETRLERDVQIGYVK 656
Qy 525 IISHPEPNDFTFYDIALLELEKPAEYSSMVRPCLPDASHVFPAGKAIWVTGHTQYG 584
Db 657 MFLEP-----YRADIALALKLSSPAITTDKHPACCPNSNMYVADRSCLYITGSGETK-G 709
Qy 585 GTGALLLQGEIRVINOCT--ENLPOOLTPPRMVCVFLSGGVDSQGGSGGLSSVEA 642
Db 710 TYGAGLKEAQLPVIENKVCNRSFLNGRSTELCAQHLAGGVDSQGGSGGLVCFEK 769

Qy 643 DGRIFQAGVSWGDCAGRNKGVYTRLEPRDWIKX 679
 Db 770 DRYILQ-GVTSWGLGCAKLTIRGVYVRVRSYSLWD 805

RESULT 23

PLBO
 plasmin (EC 3.4.21.7) precursor - bovine
 N/Alternate names: plasminogen
 C/Species: Bos primigenius taurus (cattle)
 C/Date: 30-Sep-1997 #sequence revision 28-Apr-1995 #text_change 09-Jul-2004
 C/Accession: S45046; A25835; I45961; S03736
 R/Berglund, L.; Andersen, M.D.; Petersen, T.E.
 Submitted to the EMBL Data Library, May 1994
 A/Description: Cloning and characterization of the bovine plasminogen cDNA.
 A/Reference number: S45046
 A/Accession: S45046
 A/Molecule type: mRNA
 A/Residues: 1-812

 A/Cross-references: UNIPROT:P06668; EMBL:X79402; NID:G494962; PIDD:CA55939.1; PID:G4949
 A/Experimental source: liver
 A/Note: It is uncertain whether Met-1 or Met-8 is the initiator
 R/Schaller, J.; Moser, P.W.; Danneberg-Muller, G.A.K.; Rosselet, S.J.; Kampfer, U.; Rick
 Eur. J. Biochem. 149, 267-278, 1985
 A/Title: Complete amino acid sequence of bovine plasminogen. Comparison with human plas
 A/Reference number: A25835; MUID:85203906; PMID:3846532
 A/Accession: A25835
 A/Molecule type: protein
 A/Residues: 27-334, 'D', 336-515, 'H', 517-554, 'L', 556-812 <SCH>
 R/Malinowski, D.P.; Sadler, J.E.; Davie, E.W.
 Biochemistry 23, 4243-4250, 1984
 A/Title: Characterization of a complementary deoxyribonucleic acid coding for human and
 A/Reference number: I45961; MUID:85023311; PMID:6148961
 A/Accession: I45961
 A/Status: translated from GB/EMBL/DBJ
 A/Molecule type: mRNA
 A/Residues: 706-743, 'R', 745-812 <MAL>
 A/Cross-references: GB:K02935; NID:G163551; PIDD:AAA30714.1; PID:G163552
 R/Brunsholz, R.A.; Lerch, P.G.; Schaller, J.; Rickli, E.E.; Lergler, W.; Manneberg, M.;
 Eur. J. Biochem. 114, 465-470, 1981
 A/Title: Comparison of the primary structure of the N-terminal CNBR fragments of human,
 A/Reference number: S03735; MUID:81212097; PMID:7238497
 A/Accession: S03735
 A/Molecule type: protein
 A/Residues: 27-83 <BRU>
 C/Function:
 A/Description: dissolves the fibrin of blood clots; acts as a proteolytic factor in a va
 nes the walls of the graafian follicle; also activates the urokinase-type plasminogen act
 A/Pathway: fibrinolysis
 C/Superfamily: plasmin; kringe homology; plasminogen-related protein precursor homology
 C/Keywords: duplication; fibrinolysis; glycoprotein; hydrolase; kidney; kringe; plasma;
 F:1-26/Domain: signal sequence #status predicted <SIG>
 F:8-103/Domain: plasminogen-related protein precursor homology <PLPH>
 F:27-812/Product: plasminogen #status experimental <PRO>
 F:27-103/Domain: activation peptide #status experimental <APT>
 F:104-583, 584-812/Product: plasmin #status experimental <MAT>
 F:110-588/Domain: plasmin chain A #status experimental <ACH>
 F:110-583/Domain: plasmin chain A #status experimental <ACH>
 F:192-269/Domain: kringe homology <KR1>
 F:282-359/Domain: kringe homology <KR2>
 F:384-461/Domain: kringe homology <KR4>
 F:485-564/Domain: kringe homology <KR5>
 F:584-812/Domain: plasmin chain B #status experimental <BCH>
 F:584-805/Domain: trypsin homology <TRY>
 F:56-80, 60-68, 110-186, 131-171, 159-183, 192-269, 195-333, 213-252, 241-264, 282-359, 303-342, 33
 bonds: #status predicted
 F:315/Binding site: carbohydrate (Aen) (covalent) #status experimental
 F:365/Binding site: carbohydrate (Ser) (covalent) #status experimental
 F:624, 667, 762/Active site: His, Asp, Ser #status predicted
 Query Match 12.7%; Score 481.5; DB 1; Length 812;
 Best Local Similarity 27.3%; Pred. No. 3.4e-24;

Matches 151; Conservative 73; Mismatches 220; Indels 109; Gaps 24;

Qy 181 SPYRPHYPRNIDTWNIEV-----PNNQ-----HVAVRFRFFL----- 215
 Db 310 TPKHKNRTPEPNFCK-NLEENYCRNPNGEKAPWCYTTNSRVRWEYCTIPGCESSPLSTER 368
 Qy 216 ----LEPRACGTCPKDYVEINGEKYGERSQFV--TSNSKITYRFH--SDQSYDT 266
 Db 369 MDVAVPEPQ--PVPQCHNGNGSYGTSSTITGKCCQSWSMTHRLKTPENVPNA 426
 Qy 267 GLFAEYLSYSDSPDPCQFT-----CRTERCIR--KELRCGDMACTDHSDELNCS 315
 Db 427 GLTMNYCRNPDADSPWCYTTDRVRWEFCNLKCKSETPEQVPAAPAPAEVNEPPEAD 486
 Qy 316 CDAGHOTCKNKR-----CKPLFWGDSVNDGDNBDGCGSPATPFCNSCKLCS 369
 Db 487 IGTGKSYRGKAKATTVAGVPCQ--WAAQEPHQHSIFPE-----TNPQGLER 532
 Qy 370 QQCNKDCDGDSPDEASCPRKNVVTCTKHYRCINGLCSKNPECDGKEDCDSDSDE-X 428
 Db 533 NYCNKP--GD-----VNGPWCYTNMPKRPDYC--DVPQCESSPDCGKPKYEPK 578
 Qy 429 DCDGLRSFTRQARVGTADDEGEMFQVSLHALGGHICGASLSPNMLVSAARCYID 488
 Db 579 KC-----SGRIVGCVSKPSHWQVSLRRSSR-HFCGGTLLSPKRWLTFAHCLDN 628
 Qy 489 DGRGRYDPTQWRAFLGLHQSGRSARQVGRRLKRIISHPFDFPFDDIALLEKRP 548
 Db 629 ILALSF-----YKVLGAHNEKARE-OSVQEPVSRLEFRP--SQADILALKLSRP 676
 Qy 549 AEYSMRPCLPDPASHVPAGKAIWYTGNGHYQYGGTGAIILOKEIRYINQTT--EN 606
 Db 677 AITKEVTPACLPENPMVAARTECYITGGERQ--GRFGGLKEALPIPIENKVCARNE 735
 Qy 607 LLPQITPRMVCVGLSGVDSQGSGLSPVSEADRIFQAGVSWGDCAGRNKGV 666
 Db 736 YLDGRVKTETLCAGHLIGTDSQGSGLGVLCEKDKYILQ-GVTSWGLGCAKLTIRGV 794
 Qy 667 YTRLEPRDWIKX 679
 Db 795 YVRVSPVPIER 807

RESULT 24

B30848
 plasmin (EC 3.4.21.7) precursor - rhesus macaque
 C/Species: Macaca mulatta (rhesus macaque)
 C/Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 09-Jul-2004
 C/Accession: B32869; B30848
 R/Tomlinson, J.E.; McLean, J.W.; Lamn, R.M.
 J. Biol. Chem. 264, 5957-5965, 1989
 A/Title: Rhesus monkey apolipoprotein(a). Sequence, evolution, and sites of synthesis.
 A/Reference number: A32869; MUID:89174660; PMID:2925663
 A/Accession: B32869
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-810 <TOM>
 A/Cross-references: UNIPROT:P12545; GB:J04697; NID:G342272; PIDD:AAA36901.1; PID:G342273
 C/Superfamily: plasmin; kringe homology; plasminogen-related protein precursor homology
 C/Keywords: fibrinolysis; glycoprotein; hydrolase; kidney; kringe; serine proteinase
 F:1-9/Domain: signal sequence #status predicted <SIG>
 F:103-181/Domain: kringe homology <KR1>
 F:185-262/Domain: kringe homology <KR2>
 F:275-352/Domain: kringe homology <KR3>
 F:377-454/Domain: kringe homology <KR4>
 F:481-560/Domain: kringe homology <KR5>
 F:581-803/Domain: trypsin homology <TRY>
 F:49-73, 53-61, 103-181, 124-164, 152-176, 185-262, 188-316, 206-245, 234-257, 275-352, 296-335, 324
 bonds: #status predicted
 F:622, 665, 760/Active site: His, Asp, Ser #status predicted
 Query Match 12.7%; Score 481; DB 2; Length 810;

Best Local Similarity 29.1%; Pred. No. 3.6e-24;
Matches 159; Conservative 54; Mismatches 169; Indels 164; Gaps 26;

QY 235 EKYC-----GERSOVVTSNSKITYR-----256
 DB 321 ENYRNPDEGKAPMCYTNNSQ---VRWEYCKIPBCESSEPVSTELDPAPPELTPVVOEC 377
 QY 257 FHSQ--QSUTDTGFLAEVLSYDSSDPCEQFTCRTRGCI RKELRCDGNADCTDHSDELNCS 315
 DB 378 YHGGQSYRGT-----SSTTTTGK-----KQSSMSMTPH-----407
 QY 316 CDAGHOTCKRKKFKKPLFMVCDSDVNDGDNDSDEGSCG---PQOTRCSNGKCLSKSQ 371
 DB 408 ---WHEKTPENFPAAGL-----TNNYCRNPADAGPMPCTTDPSEVRMEYCNLK-----K 453
 QY 372 CNG-----KDDC--GDGSDASCPKVVNVTC-----KHT 399
 DB 454 CSGTSGSVAAPPPVAQLPDAETPSEBDCMFGNGKGYRG--KKATTVTGTPOEAMAAQEPHS 512
 QY 400 YRCIN-----GL-----CLSK-----GNPEC--DGEKDCSDGSDKDC-----DCGLRSF 437
 DB 513 HRITPETNPRAAGEKNYCRNPDDVGGPWCYTNPNRKLFDYCDVPCQAASSFQCGKRPQV 572
 QY 438 TRQ---ARVVGSTADAGEWPMQVSLHALGCGHICGASLISPMVLVSAHCTYIDRGFRY 494
 DB 573 EPKPCPGRVVGGCVAYPHSMPWQISLRTRLGMHFCGGGLISPEWVLTAAHCL-----EKS 627
 QY 495 SDPTQMTAFGLHDOSQASAPGVERRLKRIISHPFNDFFDYDIALLEKPAEYSSM 554
 DB 628 SRPFYKYLICAH--REVALEPHVEIEVSKMFSP-----ARADILALKSSPAIITTDK 680
 QY 555 VRPILCPDASHVPAGKAIWVTGMHTQYGGTGLILQKEIRVYNQTCF--NLDPQOI 612
 DB 681 VIPACLSPPNVVADRTCEFTTGGGERTQ--GYTAGGLLKEARLPYIENKVCNRYEFLNGTV 739
 QY 613 TPRMCTGFLSGVDSQCGSGSLSSVEADGRIFQAGVYVWEGCQQRKRGYTYTLPL 672
 DB 740 KTELCAAGHLAGTDSQCGSGSLVCFEKDKYTLQ--GVTSWIGCARPNKPGYVAVSR 798
 QY 673 FRDWIK 678
 DB 799 FVTWIE 804

RESULT 25
 S33777
 Species: Rattus norvegicus (Norway rat)
 C>Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 16-Aug-2004
 C/Accession: S33777; S32013
 C/Parley, D.; Raymond, F.; Nick, H.
 Biochim. Biophys. Acta 1173, 350-352, 1993
 A>Title: Cloning and sequence analysis of rat hepsin, a cell surface serine proteinase.
 A/Reference number: S33777; MUID:93305733; PMID:8318546
 A/Accession: S33777
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-416 <FAR>
 A/Cross-references: UNIPROT:O05511; EMBL:X70900; NID:957928; PION:CAA50256.1; PID:957928
 C/Superfamily: trypsin homology
 C/Keywords: hydrolase; liver; serine proteinase; transmembrane protein
 F/22-44/Domain: transmembrane #status predicted <TNN>
 F/162-399/Domain: trypsin homology <TRY>
 F/187-203,290-358,321-337,348-380/Disulfide bonds: #status predicted
 F/202,256,352/Active site: His, Asp, Ser #status predicted

Query Match 12.7%; Score 480.5; DB 1; Length 416;
 Best Local Similarity 38.6%; Pred. No. 2e-24;
 Matches 105; Conservative 42; Mismatches 96; Indels 29; Gaps 10;

QY 431 DCGLRSTQRQARVVGTDADGEMPMQVSLHALGCGHICGASLISPMVLVSAHCTYIDR 490
 DB 151 DCGRRKLVD--RIVGGQDSSLGRPMQVSLRYDGT--HLCCGSLISGMVLTAAHCFPE-- 206

QY 491 GFRYSDPTQMTAFGLHDOSQASAPGVERRLKRIISH-----PFNDFTPD---YDIALL 543
 DB 207 --ENRVLSTRRRVPAQ---AAARTSPHAYVQGVAVIYHGGLP--FRDPTIDENSNDIALV 260
 QY 544 ELEKPAEYSSMVRPILCPDASHVPAGKAIWVTGMHTQYGGTGLILQKEIRVYNQTT 603
 DB 261 HLSSSLPRTYIOPVCLPAAGALVDGKCTVGVNGTQYGGQAVVLQEARVPIISNEV 320
 QY 604 CE--NLDPQOITPRMCMVGLSGVDSQCGSGAPL---SSVADGRIFQAGVWMDGC 658
 DB 321 CNEPDPFGNOIKRMCACAGPEGGIDACQSDSGHVPCEBRISGTRMLCGIVSNQTCG 380
 QY 659 AQRNKEGVYTRLPFRDWI-----KENTGV 683
 DB 381 ALARKPGVYTKVIDPFRWIFQAIKTHSEATGM 412

RESULT 26
 KFHUI
 coagulation factor Xla (EC 3.4.21.27) precursor [validated] - human
 N/Alternate names: antihemophilic factor C; plasma thromboplastin antecedent
 C/Species: Homo sapiens (man)
 C/Date: 13-Aug-1986 #sequence_revision 26-May-1994 #text_change 09-Jul-2004
 C/Accession: A27431; A00920; A37940
 R/Asakai, R.; Davie, E.W.; Chung, D.W.
 Biochemistry 26, 7221-7228, 1987
 A>Title: Organization of the gene for human factor XI.
 A/Reference number: A27431; MUID:88107663; PMID:2827746
 A/Accession: A27431
 A/Molecule type: DNA
 A/Residues: 1-625 <ASA>
 A/Cross-references: UNIPROT:P03951; GB:M18295
 A/Note: the sequence shown follows the authors' translation
 R/Fujikawa, K.; Chung, D.W.; Hendrickson, L.E.; Davie, E.W.
 Biochemistry 25, 2417-2424, 1986
 A>Title: Amino acid sequence of human factor XI, a blood coagulation factor with four car
 A/Reference number: A00920; MUID:86243360; PMID:3636155
 A/Accession: A00920
 A/Molecule type: mRNA
 A/Residues: 1-625 <FLU>
 A/Cross-references: GB:M13142; NID:g182832; PION:AAA52487.1; PID:g182833
 R/McMullen, B.A.; Fujikawa, K.; Davie, E.W.
 Biochemistry 30, 2056-2060, 1991
 A/Title: Location of the disulfide bonds in human coagulation factor XI: the presence of
 A/Reference number: A37940; MUID:91152017; PMID:1998667
 A/Accession: A37940
 A/Molecule type: protein
 A/Residues: 28-33;35-49,'X',51-55,'X',57-63;70-75,'X',77-79;107-109,'X',111-112;132-139,'
 ;280-282,'X',284;285-297;313-316,'X',318-319;320-326,'X',328-330,'X',347-349;373,'X',375;
 C/Comment: The proenzyme consists of two identical chains linked by one or more disulfide
 he active site, and a heavy chain, which associates with high molecular weight (HMW) kin
 C/Genetics:
 A/Gene: GDB:F11
 A/Cross-references: GDB:119891; OMIM:264900
 A/Map position: 4q35-4q35
 A/Intons: 19/1; 73/2; 109/1; 162/2; 199/1; 252/2; 289/1; 343/2; 379/1; 435/2; 494/1; 526
 C/Function:
 A>Description: catalyzes the proteolytic activation of coagulation factor IX
 A/Pathway: blood coagulation intrinsic pathway
 C/Superfamily: coagulation factor XI; trypsin homology
 C/Keywords: blood coagulation; duplication; glycoprotein; hemophilia C; homodimer; hydro
 F/1-18/Domain: signal sequence #status predicted <SIG>
 F/13-387/Product: coagulation factor Xla heavy chain #status experimental <HCH>
 F/13-108/Domain: apple repeat <AP1>
 F/109-198/Domain: apple repeat <AP2>
 F/199-288/Domain: apple repeat <AP3>
 F/290-379/Domain: apple repeat <AP4>
 F/388-625/Product: coagulation factor Xla light chain #status experimental <LCH>
 F/388-618/Domain: trypsin homology <TRY>
 F/20-103,514-581,571-599/Disulfide bonds: #status predicted
 F/22/Disulfide bonds: interchain #status experimental
 F/46-76,50-56,110-193,136-165,140-146,200-283,226-255,230-236,291-374,317-346,321-327,380

F:90,126,353,450/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:339/Disulfide bonds: interchain #status predicted (X1a) #status experimental
F:337-388/Cleavage site: Arg-116 (coagulation factor X1a) #status experimental
F:431,480,575/Active site: His, Asp, Ser #status predicted
F:491/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 12.7%; Score 480; DB 1; Length 625;

Best Local Similarity 35.7%; Pred. No. 3.3e-24; Indels 56; Gaps 13;

Matches 119; Conservative 42; Mismatches 116;

QY 382 SDEASCPKV--NVYCTKHTYRCLNGLCLSKNPECDKEDC-----SDGSDK----- 428

Db 313 SHEA--CQKLCNNAVRCQFFYTTPAQASC-----NEGKGCYKLKLSNGSPFTILHGRG 364

QY 429 -----DCCGLRSTTRQARVVGCTDADEGEPMOVSIALAG--CGHICGASLI 474

Db 365 GISGYTLRLCKMNECTTKI--KPRIVGCTASVRGEPMQVTLHTTSPTRHICGGSII 421

QY 475 SPNNLVSAAHCYIDRGRYSDFQMTAFILGHDQSO----RSAPGVQERLKRILISHPF 530

Db 422 GNMILTPAHCF-----YGVESPILKRVYSGILNQSEIKEDTSPFVQVE-----IIHDQ 471

QY 531 FNDTPDYIALLELEKPAEYSSMWPRICLPDASHVPAGKAIWVTGHTOYGTGALI 590

Db 472 YKMAESGYDIALKLETTVNTDSORPICLPKSDRNVITYDCWVTGMYKADKIQNT 531

QY 591 LQKEIRVINTGCE--NLFPQITPRMKCVGFLSGVDSCGDSGGPLSSVEADGRIFQ- 648

Db 532 LQKAKIPLVTNEBCQKRYGRHKITHKMI CAGYREGSKDCKGDSGPLSC--KINBEVWL 589

QY 649 AGVYSGDGCQARKKPGVYTRPLPRDMIKENT 681

Db 590 VGITWSGSCAQRRERPGVTVNVEYVDILKXT 622

RESULT 27

A30351

C:Species: Canis lupus familiaris (dog)

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004

C:Accession: A30351; 146201

R:Evans, J.P.; Matzke, H.H.; Ware, J.L.; Stafford, D.W.; High, K.A.

A:Title: Molecular cloning of a cDNA encoding canine factor IX.

A:Reference number: A30351; MUID:89323338; PMID:2752110

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-452 <EVA>

A:Cross-references: UNIPROT:P19540; GB:M21757; NID:g972719; PIDN:AAA75006.1; PID:g163948

R:Axelrod, J.H.; Read, W.S.; Brinkhaus, K.M.; Verma, I.M.

Proc. Natl. Acad. Sci. U.S.A. 87, 5173-5177, 1990

A:Title: Phenotypic correction of factor IX deficiency in skin fibroblasts of hemophilic

A:Reference number: 146201; MUID:90311364; PMID:2367529

A:Accession: 146201

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-452 <AEB>

A:Cross-references: GB:M33826; NID:g163949; PIDN:AAA30844.1; PID:g163950

C:Superfamily: coagulation factor X; EGF homology; Gla domain homology; trypsin homology

C:Keywords: beta-hydroxyaspartic acid, blood coagulation, calcium binding, carboxylglutamate

F:1-21/Domain: signal sequence #status predicted <SIG>

F:22-40/Domain: propeptide #status predicted <SIG>

F:24-84/Domain: Gla domain homology <Gla>

F:41-452/Product: coagulation factor IX #status predicted <MAT>

F:90-121/Domain: EGF homology <EGF>

F:127-163/Domain: EGF homology <EGF>

F:218-445/Domain: trypsin homology <TRY>

F:46,47,54,56,59,60,65,66,69,72,75,79/Modified site: gamma-carboxyglutamic acid (Glu) #

F:57-62,90-101,95-110,112-121,127-138,134-148,150-163,171-326,243-259,373-387,398-426/D1

F:258,306,402/Active site: His, Asp, Ser #status predicted

Query Match 12.5%; Score 474.5; DB 1; Length 452;

Best Local Similarity 28.4%; Pred. No. 5.4e-24;

Matches 143; Conservative 64; Mismatches 172; Indels 125; Gaps 19;

QY 215 LLEPRRACGCPDYINEKTC--GERSQF---VTSNKNKTTFPHSDQSTTDYDGL 269

Db 34 LSRPKRYSNGKLEEFVAGNLERECIEKCSFEAREVFENTEK-----TTEFW 81

QY 270 AEVLSTVSSDPCEQFCRTGRCIRKELRGDGMADCDHSDDELNCSGDAGHPTCKKFC 329

Db 82 KQYVGDQ-----CESNPCLN---DG--VCQDINSYECWCRAGP----- 116

QY 330 KPLFWCDVYNDGSDNDEQSCCPAQ--TERCSNGKCLSKSQOQNGKDDGSDSDEASCP 388

Db 117 -----EGKNCLEDTVCTNKNRC-----KQFCGLGPD----- 143

QY 389 KQNVYCTKHTYRCLNGLCLSKNPECD-----GKEDCS 422

Db 144 --NKVYCS-----CTTGQQLAEQDQSCPEAVPPFCGRVSPHISMTRATETLPSNNDYE 196

QY 423 DGSD--EKDCGGLRSFTROARVVGCTDADEGEPMOVSIALAGGHTCGASLISPNLVLS 481

Db 197 NSTEVEKILNVTPQLNDFRVRVGGKDAKPGQFPWQVLLNGKVA--FCGGSILNEKVVY 255

QY 482 AAHCYIDRGRYSDFQMTAFILGHDQSORSAAGVQERLKRILISHPFNDP--TFDYD 539

Db 256 AAHCIEPD-----VKITIVAGEHTEKREHT--EQKRVIRITLHSHYNATINKYNHD 306

QY 540 IALLEKPAEYSSMWPRICLPDA--SHVPAGKAIWVTGHTOYGTGALI LQKEIR 597

Db 307 IALLEDEPPLNSVYPICTIADREYSNIFLRKSGSVSGMGRFNNGRASISIIQYLVKP 366

QY 598 VINQTCENILPQITPRMKCVGFLSGVDSCGDSGGPLSSVEADGRIFQAGVSGDG 657

Db 367 LVDPATCLRSKTKFTTNNMFCAGHSGKDSGDSGCP--HTVEVGIISFLTGIIISGEE 425

QY 658 CAQRNKPQVYTRPLPRDMIKENT 681

Db 426 CAMKKGITTKYSRYVNWIKENT 449

RESULT 28

A59271

R:Reactive factor (EC 3.4.21.-) 2 precursor - human

N:Alternate names: mannose binding protein-2 associated serine proteinase 2 (MASP-2)

C:Species: Homo sapiens (man)

C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 09-Jul-2004

C:Accession: A59271

R:Thiel, S.; Vorup-Jensen, T.; Stover, C.M.; Schwaeble, W.J.; Laursen, S.B.; Poulsen, K.

Nature 386, 506-510, 1997

A:Title: A second serine protease associated with mannan-binding lectin that activates co

A:Reference number: A59271; MUID:97242412; PMID:9087411

A:Accession: A59271

A:Status: nucleic acid sequence not shown; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-686 <JEN>

A:Cross-references: UNIPROT:O00187; GB:Y09926; NID:g4007626; PIDN:CAA71059.1; PID:g4007626

A:Experimental source: tissue liver

A>Note: submitted to GenBank, December 1996

A:Note: parts of this sequence, including the amino end of the mature protein, were deter

C:Genetics: GDB:MASP2

A:Cross-references: GDB:6071500

A:Map position: 1p36.2-1p36.3

C:Superfamily: complement activating serine proteases C1r/C1s/MASP, C1r/C1s repeat homolo

C:Keywords: beta-hydroxyaspartate; complement pathway; duplication; hydrolase; serine pr

F:1-15/Domain: signal sequence #status predicted <SIG>

F:16-44,445-686/Product: Ra-reactive factor 2 #status predicted <MAT>

F:119-134/Domain: C1r/C1s repeat homology <C1R1>

F:142-180/Domain: EGF homology <EGF>

F:184-223/Domain: C1r/C1s repeat homology <C1R2>

F:300-351/Domain: complement factor H repeat homology <FH1>

F:366-430/Domain: complement factor H repeat homology <FH2>

F:445-679/Domain: trypsin homology <TRY>


```

Db      57  KQGSWSSWTPH-----MHQKTPREKYPNADL-----TMNYCRNPDGDKGPMCYTTD 101
Qy      354 PAQTFR-CSNGKCLSKSQQCNKDDC---GDSPDASCPRKNNVTCT-----KHTY 400
Db      102 PSVMWEPFNLKRC-SETVQEPSEPDGCMIGIKGYGKKATTVTGTRCQAMAAQEPHRSI 160
Qy      401 -----RC-----LNGLCLSKNPB-----CDQKE-----DCSDGSDDE-K 428
Db      161 FTPEANPWANLEKRYCERNPDGDVNGPMCYTMNPKLPDYCDVPQCESSPFDCGKRYEPK 220
Qy      429 DCDGCLRSFTQARVVGSTDADEGEWPQVSLHALGQGHICGASLISPMLVSAHCYID 488
Db      221 KC-----SGRIVGGCVAILAHSMFWQISLRTFRGRHFCGTLISPFWTLTAHCL-- 269
Qy      489 DRGRYSDPTQWTAFLGLHDQSORSAPGVQERRLKRIISHPFENDFTFDYDIALLEKXP 548
Db      270 ---ERSSRPSTYKVLGTNHE-LRLAAGAOQIDVSKLFLER-----SRADIALKLSSP 319
Qy      549 ABEYSMWPRICLPDASHVFPAGKAIWYTGWGHTOYCGTALILQKEIRVINQTTCE--N 606
Db      320 AITQWYIPACLPADYVAVANMAECFTVGMGETQ-DSSNAGVLKEAQLPVLENKVCNRYE 378
Qy      607 LLPQOITPRMMCVGFLSGVDSCQDSCGPLSVADGRIFOAGVSWGDCQOBNKPGV 666
Db      379 YLNGRVXSTELCAGHLVGVVSCQDSCGPLVCFEKDKYILQ-GVTSWGLGACARNKPGV 437
Qy      667 YTRLPLFRDWIK 678
Db      438 YVRVSSPIMIE 449

```

Search completed: November 29, 2004, 08:33:22
 Job time : 30.3127 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 29, 2004, 08:32:39 ; Search time 88.8166 Seconds

(without alignments)
2727.603 Million cell updates/sec

Title: US-09-936-333-5

Perfect score: 3789

Sequence: 1 MABERVLMPLPRARSLKSFV.....PGVYTRLPFRMDIKENTGV 683

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1575965 seqs, 354694765 residues

Total number of hits satisfying chosen parameters: 1575965

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : Published Applications AA:*

1: /cgn2_6/prodata/2/pubppaa/US07_PUBCOMB.pep:*
2: /cgn2_6/prodata/2/pubppaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/prodata/2/pubppaa/US06_NEW_PUB.pep:*
4: /cgn2_6/prodata/2/pubppaa/US07_NEW_PUB.pep:*
5: /cgn2_6/prodata/2/pubppaa/PCTUS_PUBCOMB.pep:*
6: /cgn2_6/prodata/2/pubppaa/US08_NEW_PUB.pep:*
7: /cgn2_6/prodata/2/pubppaa/US08_PUBCOMB.pep:*
8: /cgn2_6/prodata/2/pubppaa/US09_PUBCOMB.pep:*
9: /cgn2_6/prodata/2/pubppaa/US09B_PUBCOMB.pep:*
10: /cgn2_6/prodata/2/pubppaa/US09C_PUBCOMB.pep:*
11: /cgn2_6/prodata/2/pubppaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/prodata/2/pubppaa/US09C_PUBCOMB.pep:*
13: /cgn2_6/prodata/2/pubppaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/prodata/2/pubppaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/prodata/2/pubppaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/prodata/2/pubppaa/US10D_PUBCOMB.pep:*
17: /cgn2_6/prodata/2/pubppaa/US10D_NEW_PUB.pep:*
18: /cgn2_6/prodata/2/pubppaa/US10D_NEW_PUB.pep:*
19: /cgn2_6/prodata/2/pubppaa/US60_NEW_PUB.pep:*
20: /cgn2_6/prodata/2/pubppaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3760	99.2	762	US-10-729-807-1	Sequence 1, Appl1
2	3760	99.2	855	US-10-295-027-1185	Sequence 1185, Ap
3	3760	99.2	855	US-10-072-012-353	Sequence 353, App
4	3760	99.2	855	US-10-072-012-412	Sequence 412, App
5	3760	99.2	855	US-10-072-012-419	Sequence 419, App
6	3756	99.1	855	US-10-072-012-354	Sequence 354, App
7	3756	99.1	855	US-10-072-012-420	Sequence 420, App
8	3756	99.1	855	US-10-037-417-132	Sequence 132, App
9	3754	99.1	757	US-10-072-012-44	Sequence 44, Appl
10	3754	99.1	855	US-09-776-191-2	Sequence 2, Appl1
11	3754	99.1	855	US-10-099-700A-2	Sequence 2, Appl1
12	3754	99.1	855	US-10-190-030B-2	Sequence 2, Appl1
13	3754	99.1	855	US-10-302-840A-2	Sequence 2, Appl1

14	3754	99.1	855	US-10-267-219-2	Sequence 2, Appl1
15	3754	99.1	855	US-10-112-221A-2	Sequence 2, Appl1
16	3754	99.1	855	US-10-104-271-2	Sequence 2, Appl1
17	3754	99.1	855	US-10-147-211A-2	Sequence 2, Appl1
18	3754	99.1	855	US-10-156-214A-2	Sequence 2, Appl1
19	3754	99.1	855	US-10-072-012-352	Sequence 352, App
20	3754	99.1	855	US-10-072-012-411	Sequence 411, App
21	3754	99.1	855	US-10-072-012-418	Sequence 418, App
22	3754	99.1	855	US-10-600-187-2	Sequence 2, Appl1
23	3745	98.8	851	US-10-276-774-1798	Sequence 1798, Ap
24	3745	98.8	851	US-10-296-115-1143	Sequence 1143, Ap
25	3259.5	86.0	782	US-10-097-340-312	Sequence 312, App
26	3209	84.7	852	US-09-900-751-2	Sequence 2, Appl1
27	3209	84.7	855	US-10-072-012-355	Sequence 355, App
28	3209	84.7	855	US-10-072-012-413	Sequence 413, App
29	3197	84.4	855	US-10-072-012-356	Sequence 356, App
30	3197	84.4	855	US-10-072-012-414	Sequence 414, App
31	3197	84.4	855	US-10-072-012-417	Sequence 417, App
32	3123	82.4	902	US-10-333-743-3	Sequence 3, Appl1
33	3123	82.4	902	US-10-600-187-10	Sequence 10, Appl
34	3123	82.4	902	US-10-297-987B-11	Sequence 11, Appl
35	2797	73.8	620	US-09-925-301-1193	Sequence 1193, Ap
36	2254.5	59.5	845	US-10-072-012-415	Sequence 415, App
37	1319	34.8	241	US-09-776-191-50	Sequence 50, Appl
38	1319	34.8	241	US-10-099-700A-4	Sequence 4, Appl1
39	1319	34.8	241	US-10-092-004A-2	Sequence 2, Appl1
40	1319	34.8	241	US-10-190-030B-4	Sequence 4, Appl1
41	1319	34.8	241	US-10-302-840A-4	Sequence 4, Appl1
42	1319	34.8	241	US-10-267-219-4	Sequence 4, Appl1
43	1319	34.8	241	US-10-112-221A-4	Sequence 4, Appl1
44	1319	34.8	241	US-10-104-271-4	Sequence 4, Appl1
45	1319	34.8	241	US-10-147-211A-4	Sequence 4, Appl1
46	1029	27.2	850	US-09-888-615-108	Sequence 108, App
47	1019.5	26.9	833	US-10-333-743-7	Sequence 7, Appl1
48	1014	26.8	799	US-10-072-012-410	Sequence 410, App
49	1014	26.8	799	US-10-072-012-416	Sequence 416, App
50	1005	26.5	802	US-09-888-615-113	Sequence 113, App
51	1005	26.5	802	US-09-978-295A-169	Sequence 169, App
52	1005	26.5	802	US-09-978-697-169	Sequence 169, App
53	1005	26.5	802	US-09-978-192A-169	Sequence 169, App
54	1005	26.5	802	US-09-999-832A-169	Sequence 169, App
55	1005	26.5	802	US-09-978-189-169	Sequence 169, App
56	1005	26.5	802	US-09-978-608A-169	Sequence 169, App
57	1005	26.5	802	US-09-978-585A-169	Sequence 169, App
58	1005	26.5	802	US-09-978-191A-169	Sequence 169, App
59	1005	26.5	802	US-09-978-403A-169	Sequence 169, App
60	1005	26.5	802	US-09-978-564A-169	Sequence 169, App
61	1005	26.5	802	US-09-999-833A-169	Sequence 169, App
62	1005	26.5	802	US-09-981-915A-169	Sequence 169, App
63	1005	26.5	802	US-09-978-824-169	Sequence 169, App
64	1005	26.5	802	US-09-978-585A-169	Sequence 169, App
65	1005	26.5	802	US-09-999-834A-169	Sequence 169, App
66	1005	26.5	802	US-09-978-423A-169	Sequence 169, App
67	1005	26.5	802	US-09-978-193A-169	Sequence 169, App
68	1005	26.5	802	US-09-999-830A-169	Sequence 169, App
69	1005	26.5	802	US-09-978-757A-169	Sequence 169, App
70	1005	26.5	802	US-09-978-187B-169	Sequence 169, App
71	1005	26.5	802	US-09-978-643A-169	Sequence 169, App
72	1005	26.5	802	US-09-978-375A-169	Sequence 169, App
73	1005	26.5	802	US-09-978-298A-169	Sequence 169, App
74	1005	26.5	802	US-09-978-188A-169	Sequence 169, App
75	1005	26.5	802	US-09-978-681A-169	Sequence 169, App
76	1005	26.5	802	US-09-978-194A-169	Sequence 169, App
77	1005	26.5	802	US-09-999-828A-169	Sequence 169, App
78	1005	26.5	802	US-09-978-544A-169	Sequence 169, App
79	1005	26.5	802	US-09-978-665A-169	Sequence 169, App
80	1005	26.5	802	US-09-978-802A-169	Sequence 169, App
81	1005	26.5	802	US-09-999-831A-169	Sequence 169, App
82	1005	26.5	802	US-10-017-081A-169	Sequence 169, App
83	1005	26.5	802	US-10-167-749-169	Sequence 169, App
84	1005	26.5	802	US-10-013-921A-169	Sequence 169, App
85	1005	26.5	802	US-10-013-929A-169	Sequence 169, App
86	1005	26.5	802	US-10-013-929A-169	Sequence 169, App

```
87 1005 26.5 802 14 US-10-016-177A-169 Sequence 169, App
88 1005 26.5 802 14 US-10-166-709A-169 Sequence 169, App
89 1005 26.5 802 14 US-10-143-031A-169 Sequence 169, App
90 1005 26.5 802 14 US-10-143-030A-169 Sequence 169, App
91 1005 26.5 802 14 US-10-002-967A-169 Sequence 169, App
92 1005 26.5 802 14 US-10-017-083A-169 Sequence 169, App
93 1005 26.5 802 14 US-10-145-128A-169 Sequence 169, App
94 1005 26.5 802 14 US-10-017-191A-169 Sequence 169, App
95 1005 26.5 802 14 US-10-143-028A-169 Sequence 169, App
96 1005 26.5 802 14 US-10-143-029A-169 Sequence 169, App
97 1005 26.5 802 14 US-10-145-089A-169 Sequence 169, App
98 1005 26.5 802 14 US-10-165-067A-169 Sequence 169, App
99 1005 26.5 802 14 US-10-145-017A-169 Sequence 169, App
100 1005 26.5 802 14 US-10-164-728A-169 Sequence 169, App
```

ALIGNMENTS

```
RESULT 1
US-10-729-807-1
; Sequence 1, Application US/10729807
; Publication No. US20040132158A1
; GENERAL INFORMATION:
; APPLICANT: BANDMAN, Olga; HILLMAN, Jennifer L.
; APPLICANT: TANG, Y. Tom; LAL, Preeti G.
; APPLICANT: YUE, Henry; AZIMZAI, Valda
; APPLICANT: BAUGHN, Mariah R.; LU, Dyung Aina M.
; TITLE OF INVENTION: HUMAN PEPTIDASES
; FILE REFERENCE: PF-0651-1 DIV
; CURRENT APPLICATION NUMBER: US/10/729, 807
; CURRENT FILING DATE: 2003-12-05
; PRIOR APPLICATION NUMBER: US 09/889, 238
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: PCT/US00/00641
; PRIOR FILING DATE: 2000-01-11
; PRIOR APPLICATION NUMBER: US 60/172, 247
; PRIOR FILING DATE: 1999-01-11
; PRIOR APPLICATION NUMBER: US 60/132, 253
; PRIOR FILING DATE: 1999-05-03
; PRIOR APPLICATION NUMBER: US 60/136, 653
; PRIOR FILING DATE: 1999-05-27
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PERL Program
; SEQ ID NO 1
; LENGTH: 762
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: 155179CD1
US-10-729-807-1
```

```
Query Match 99.2%; Score 3760; DB 16; Length 762;
Best Local Similarity 99.4%; Pred. No. 2e-275;
Matches 679; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 1 MAERVVMLPPRRSLKSFVVTGVVAFPTDSTKYORTODNSCGFGLHARGVEMRFTTPG 60
DB 80 MAERVVMLPPRRSLKSFVVTGVVAFPTDSTKYORTODNSCGFGLHARGVEMRFTTPG 139
QY 61 FPDSPYAHARCOMALRGDADSVLSLTFRSFDLASCDERSGDLVTVNTLSPMEPHALVQ 120
DB 140 FPDSPYAHARCOMALRGDADSVLSLTFRSFDLASCDERSGDLVTVNTLSPMEPHALVQ 199
QY 121 LCGTTPPSYMLTTHSSQNVLLITLITERRHPFEATFQQLPRMSSCGRLKRAQGTEN 180
DB 200 LCGTTPPSYMLTTHSSQNVLLITLITERRHPFEATFQQLPRMSSCGRLKRAQGTEN 259
QY 181 SPYYPGHPNICTNWIEVPPNNOHVVRKFPFLLEPRRACGTCPCDYYVEINKEKXCGE 240
DB 260 SPYYPGHPNICTNWIEVPPNNOHVVRKFPFLLEPRRACGTCPCDYYVEINKEKXCGE 319
```

```
QY 241 RSQFVVTNSNKTITVRFHSDOSYTDGFLAFLYLSYDSDPCQGFCTCRTRCIRKELRCD 300
DB 320 RSQFVVTNSNKTITVRFHSDOSYTDGFLAFLYLSYDSDPCQGFCTCRTRCIRKELRCD 379
QY 301 GMAICTDHSDELNCSDAGHQFTCKNFKCKPLFWVCDSVNDCDNDDEQCSCPAQTFRC 360
DB 380 GMAICTDHSDELNCSDAGHQFTCKNFKCKPLFWVCDSVNDCDNDDEQCSCPAQTFRC 439
QY 361 SNKCLSKSQCGKXGKDCGSDGDEASCPRKNVYTCRHTYRCLNGCLSKGNECDGKED 420
DB 440 SNKCLSKSQCGKXGKDCGSDGDEASCPRKNVYTCRHTYRCLNGCLSKGNECDGKED 499
QY 421 CSDGSEKDCDCGLRSFTTRQARVVGSTDADEGEMPVQVSLHALGQGHICGASLISPMVLV 480
DB 500 CSDGSEKDCDCGLRSFTTRQARVVGSTDADEGEMPVQVSLHALGQGHICGASLISPMVLV 559
QY 481 SAACHYIDRGRFYSDPTQWTAFLGLHDOSQRAAPGVQERRLRITISHPFNDFTDYDI 540
DB 560 SAACHYIDRGRFYSDPTQWTAFLGLHDOSQRAAPGVQERRLRITISHPFNDFTDYDI 619
QY 541 ALLELEKPAFYSMWBPICLPDASHVFPAGKAIWVTGMGHTGYGTAIILQKGEIRVIN 600
DB 620 ALLELEKPAFYSMWBPICLPDASHVFPAGKAIWVTGMGHTGYGTAIILQKGEIRVIN 679
QY 601 QTTCEMLLPQOITPRMVCVGLSGVDSCGDSGGLSSVEADGRIFOAGVSWGDCAQ 660
DB 680 QTTCEMLLPQOITPRMVCVGLSGVDSCGDSGGLSSVEADGRIFOAGVSWGDCAQ 729
QY 661 RNRPGYTRLPFLFRDWIKENTGV 683
DB 740 RNRPGYTRLPFLFRDWIKENTGV 762
```

```
RESULT 2
US-10-295-027-1185
; Sequence 1185, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natsaha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynn, Richard
; APPLICANT: Hevez, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Bos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer; Compositions and
; FILE REFERENCE: 018501-012500US
; CURRENT APPLICATION NUMBER: US/10/295, 027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663, 733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350, 666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335, 394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332, 464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334, 393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/347, 349
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/355, 250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356, 714
; PRIOR FILING DATE: 2002-02-13
```



```

; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: Patentn Ver. 2.1
; SEQ ID NO 1185
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-295-027-1185

Query Match      99.2%; Score 3760; DB 14; Length 855;
Best Local Similarity 99.4%; Pred. No. 2.2e-275;
Matches 679; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 MAERVVMLPPRASLSKSFVVTSVVAPPTDSTKYQRTODNSCSFGLHARGVELMRFTTPG 60
Db 173 MAERVVMLPPRASLSKSFVVTSVVAPPTDSTKYQRTODNSCSFGLHARGVELMRFTTPG 232
Qy 61 FPDSPYPAHARCOMALRGDADSVSLTFRSPDLASCDERSGDLVTYVNTLSPMEPHALVQ 120
Db 233 FPDSPYPAHARCOMALRGDADSVSLTFRSPDLASCDERSGDLVTYVNTLSPMEPHALVQ 292
Qy 121 LCGTTPPSYNLTFRSSQNVLLITLITNTERRHPEFATFPQLPRMSSCGRLRAOQTFN 180
Db 293 LCGTTPPSYNLTFRSSQNVLLITLITNTERRHPEFATFPQLPRMSSCGRLRAOQTFN 352
Qy 181 SPYRGHYPPNIDCTNMIIEVNNQHVVRKFKFYLLERPRACGTCPCDYIINEKTCGE 240
Db 353 SPYRGHYPPNIDCTNMIIEVNNQHVVRKFKFYLLERPRACGTCPCDYIINEKTCGE 412
Qy 241 RSQFVVTNSNKTIVRFHSDQSYDTDTGFLAEVLSYDSDPCPGQFTCRGRCIRKEIRCD 300
Db 413 RSQFVVTNSNKTIVRFHSDQSYDTDTGFLAEVLSYDSDPCPGQFTCRGRCIRKEIRCD 472
Qy 301 GMADCTHSDBLNCSCDAGHQTCKNKECKPLFWVCDSVNDCGNSDEQSCSPAQTFRC 360
Db 473 GMADCTHSDBLNCSCDAGHQTCKNKECKPLFWVCDSVNDCGNSDEQSCSPAQTFRC 532
Qy 361 SNGCCLSKSQOCCNGKDDCGSDSDASCPKNNVVTCTKRTYCLNGLCLSKNPNCDGED 420
Db 533 SNGCCLSKSQOCCNGKDDCGSDSDASCPKNNVVTCTKRTYCLNGLCLSKNPNCDGED 592
Qy 421 CSDSDDEKDCDCCGRLSFTROARVVGTDADGEMPMQVSLALQGHICGSLISPMNLV 480
Db 593 CSDSDDEKDCDCCGRLSFTROARVVGTDADGEMPMQVSLALQGHICGSLISPMNLV 652
Qy 481 SAACHYIDDRGRYSDDPTQWTAFLGLHDQSORSAAPGVQERLKIISHPFNDFPDYDI 540
Db 653 SAACHYIDDRGRYSDDPTQWTAFLGLHDQSORSAAPGVQERLKIISHPFNDFPDYDI 712
Qy 541 ALLELEKPAEYSSNVRPCLPDASHVFPAGKAIWVTGHTQYGGTGLLIQKBEIRYIN 600
Db 713 ALLELEKPAEYSSNVRPCLPDASHVFPAGKAIWVTGHTQYGGTGLLIQKBEIRYIN 772
Qy 601 QTTGEBNLLPQOITPRMNCVGLSGGVSDSCGDSGGPLSSVADGRIFQAGVNSGDCAQ 660
Db 773 QTTGEBNLLPQOITPRMNCVGLSGGVSDSCGDSGGPLSSVADGRIFQAGVNSGDCAQ 832
Qy 661 RNKPGVYTRPLPFRDWIKENTGV 683
Db 833 RNKPGVYTRPLPFRDWIKENTGV 855

RESULT 3
US-10-072-012-353
; Sequence 353, Application US/10072012
; Publication No. US20040033493A1
; GENERAL INFORMATION:
; APPLICANT: Tcheurnev, Velizar
; APPLICANT: Spytek, Kimberly
; APPLICANT: Zeehuusen, Bryan
; APPLICANT: Paturajan, Meera
; APPLICANT: Shimkete, Richard
; APPLICANT: Li, Li

; APPLICANT: Gangoli, Babha
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Anderson, David W.
; APPLICANT: Rastelli, Luca
; APPLICANT: Miller, Charles E.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Taulier Jr, Raymond J.
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Colman, Steven D.
; APPLICANT: Wolenc, Adam R.
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Grose, William M.
; APPLICANT: Albrook II, John P.
; APPLICANT: Lepley, Denise M.
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Burgess, Catherine E.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-258
; CURRENT APPLICATION NUMBER: US/10/072,012
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 60/265,102
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/265,514
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,412
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,395
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/266,406
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/266,767
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/267,057
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/266,975
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/267,459
; PRIOR FILING DATE: 2001-02-08
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1391
; SOFTWARE: Patentn Ver. 2.1
; SEQ ID NO 353
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-072-012-353

Query Match      99.2%; Score 3760; DB 15; Length 855;
Best Local Similarity 99.4%; Pred. No. 2.2e-275;
Matches 679; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 MAERVVMLPPRASLSKSFVVTSVVAPPTDSTKYQRTODNSCSFGLHARGVELMRFTTPG 60
Db 173 MAERVVMLPPRASLSKSFVVTSVVAPPTDSTKYQRTODNSCSFGLHARGVELMRFTTPG 232
Qy 61 FPDSPYPAHARCOMALRGDADSVSLTFRSPDLASCDERSGDLVTYVNTLSPMEPHALVQ 120
Db 233 FPDSPYPAHARCOMALRGDADSVSLTFRSPDLASCDERSGDLVTYVNTLSPMEPHALVQ 292
Qy 121 LCGTTPPSYNLTFRSSQNVLLITLITNTERRHPEFATFPQLPRMSSCGRLRAOQTFN 180
Db 293 LCGTTPPSYNLTFRSSQNVLLITLITNTERRHPEFATFPQLPRMSSCGRLRAOQTFN 352
Qy 181 SPYRGHYPPNIDCTNMIIEVNNQHVVRKFKFYLLERPRACGTCPCDYIINEKTCGE 240
Db 353 SPYRGHYPPNIDCTNMIIEVNNQHVVRKFKFYLLERPRACGTCPCDYIINEKTCGE 412
Qy 241 RSQFVVTNSNKTIVRFHSDQSYDTDTGFLAEVLSYDSDPCPGQFTCRGRCIRKEIRCD 300
Db 413 RSQFVVTNSNKTIVRFHSDQSYDTDTGFLAEVLSYDSDPCPGQFTCRGRCIRKEIRCD 472
```

```
QY 301 GMADCTDHSDELNCS CDAGHQFTCKNFKCKPLFWVCDSVNDCCGNSDEGSCSPAQTFRFC 360
D 473 GMADCTDHSDELNCS CDAGHQFTCKNFKCKPLFWVCDSVNDCCGNSDEGSCSPAQTFRFC 532
QY 361 SNGKCLSKSQOQCKGKDDCGSDGSDASCPKXNVVTCYKHYRCLNGLCLSKGNPECDKED 420
D 533 SNGKCLSKSQOQCKGKDDCGSDGSDASCPKXNVVTCYKHYRCLNGLCLSKGNPECDKED 592
QY 421 CSDGSDKDCDCGLRSFTROARVVGTTDADGEGMPQVSLHALGQGHICGASLISPNMLV 480
D 593 CSDGSDKDCDCGLRSFTROARVVGTTDADGEGMPQVSLHALGQGHICGASLISPNMLV 652
QY 481 SAACHYIDRGFRYSPTQWTAFLGLHDQSRSAPGVQERRLKRIISHPEFNDFPDYDI 540
D 653 SAACHYIDRGFRYSPTQWTAFLGLHDQSRSAPGVQERRLKRIISHPEFNDFPDYDI 712
QY 541 ALLELEKPAEYSSMWVRICLPDASHVFPAGKAIWVTGHTQYCGTGALLIQKEIRVIN 600
D 713 ALLELEKPAEYSSMWVRICLPDASHVFPAGKAIWVTGHTQYCGTGALLIQKEIRVIN 772
QY 601 QTTCEMLPQOITPRMVCVGLSGVDS CGDSCGPLSVSEADGRIFQAGVSWGDCAQ 660
D 773 QTTCEMLPQOITPRMVCVGLSGVDS CGDSCGPLSVSEADGRIFQAGVSWGDCAQ 832
QY 661 RNKPGVYTRLPPLFRDWIKENTGV 683
D 833 RNKPGVYTRLPPLFRDWIKENTGV 855

RESULT 4
US-10-072-012-412
; Sequence 412, Application US/10072012
; Publication No. US20040033493A1
; GENERAL INFORMATION:
; APPLICANT: Tchernyev, Velizar
; APPLICANT: Spytek, Kimberly
; APPLICANT: Zernusen, Bryan
; APPLICANT: Paturajan, Meera
; APPLICANT: Shimkets, Richard
; APPLICANT: Li, Li
; APPLICANT: Gangoli, Esha
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Anderson, David W.
; APPLICANT: Rastelli, Luca
; APPLICANT: Miller, Charles E.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Gusev, Vladimiro Y.
; APPLICANT: Coleman, Steven D.
; APPLICANT: Wolenc, Adam R.
; APPLICANT: Pena, Carol E. A
; APPLICANT: Putrak, Katarzyna
; APPLICANT: Groose, William M.
; APPLICANT: Alsobrook II, John P.
; APPLICANT: Lepley, Denise M.
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Burgess, Catherine E.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-258
; CURRENT APPLICATION NUMBER: US/10/072,012
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 60/265,102
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/265,514
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,412
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,395
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/266,406
```

```
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/266,767
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/267,057
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/266,975
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/267,459
; PRIOR FILING DATE: 2001-02-08
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1391
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 412
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-072-012-412
```

```
Query Match 99.4%; Score 3760; DB 15; Length 855;
Best Local Similarity 99.4%; Pred. No. 2,2e-275;
Matches 679; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 1 MAERVVMLEPPRARSLSKSPVYVSVAFPPTDSKTORTONDSQSGFLHARGVELMRFTTPG 60
D 173 MAERVVMLEPPRARSLSKSPVYVSVAFPPTDSKTORTONDSQSGFLHARGVELMRFTTPG 232
QY 61 FPDSPYPAHARCOMALRGDADSVLSLTPRSFDLASCDERGSDLVTVYNTLSPMHPALVQ 120
D 233 FPDSPYPAHARCOMALRGDADSVLSLTPRSFDLASCDERGSDLVTVYNTLSPMHPALVQ 292
QY 121 LCCTTPSPYNTLTHSSQNTLTLITNTERRHGFATFPQLPRMSSCGRLPKAGCTN 180
D 293 LCCTTPSPYNTLTHSSQNTLTLITNTERRHGFATFPQLPRMSSCGRLPKAGCTN 352
QY 181 SPYYPGHYPNIDCTNNIEVPPNNOHYKVRKFPYLLPERRACSTCKDYIENGEXYCE 240
D 353 SPYYPGHYPNIDCTNNIEVPPNNOHYKVRKFPYLLPERRACSTCKDYIENGEXYCE 412
QY 241 RSQFVVTNSNKTITVFHSDOSYTDGTFLAELYSYDSDPCPGQFTCKTGRCIRKELCD 300
D 413 RSQFVVTNSNKTITVFHSDOSYTDGTFLAELYSYDSDPCPGQFTCKTGRCIRKELCD 472
QY 301 GMADCTDHSDELNCS CDAGHQFTCKNFKCKPLFWVCDSVNDCCGNSDEGSCSPAQTFRFC 360
D 473 GMADCTDHSDELNCS CDAGHQFTCKNFKCKPLFWVCDSVNDCCGNSDEGSCSPAQTFRFC 532
QY 361 SNGKCLSKSQOQCKGKDDCGSDGSDASCPKXNVVTCYKHYRCLNGLCLSKGNPECDKED 420
D 533 SNGKCLSKSQOQCKGKDDCGSDGSDASCPKXNVVTCYKHYRCLNGLCLSKGNPECDKED 592
QY 421 CSDGSDKDCDCGLRSFTROARVVGTTDADGEGMPQVSLHALGQGHICGASLISPNMLV 480
D 593 CSDGSDKDCDCGLRSFTROARVVGTTDADGEGMPQVSLHALGQGHICGASLISPNMLV 652
QY 481 SAACHYIDRGFRYSPTQWTAFLGLHDQSRSAPGVQERRLKRIISHPEFNDFPDYDI 540
D 653 SAACHYIDRGFRYSPTQWTAFLGLHDQSRSAPGVQERRLKRIISHPEFNDFPDYDI 712
QY 541 ALLELEKPAEYSSMWVRICLPDASHVFPAGKAIWVTGHTQYCGTGALLIQKEIRVIN 600
D 713 ALLELEKPAEYSSMWVRICLPDASHVFPAGKAIWVTGHTQYCGTGALLIQKEIRVIN 772
QY 601 QTTCEMLPQOITPRMVCVGLSGVDS CGDSCGPLSVSEADGRIFQAGVSWGDCAQ 660
D 773 QTTCEMLPQOITPRMVCVGLSGVDS CGDSCGPLSVSEADGRIFQAGVSWGDCAQ 832
QY 661 RNKPGVYTRLPPLFRDWIKENTGV 683
D 833 RNKPGVYTRLPPLFRDWIKENTGV 855
```

```
RESULT 5
US-10-072-012-419
```

```
; Sequence 419, Application US/10072012
; Publication No. US20040033493A1
; GENERAL INFORMATION:
; APPLICANT: Tchernev, Velizar
; APPLICANT: Spytek, Kimberly
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Patuturajan, Meera
; APPLICANT: Shimkets, Richard
; APPLICANT: Li, Li
; APPLICANT: Gangoli, Beha
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Anderson, David W.
; APPLICANT: Rastelli, Luca
; APPLICANT: Miller, Charles E.
; APPLICANT: Gerlach, Valerie E.
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Colman, Steven D.
; APPLICANT: Wolenc, Adam R.
; APPLICANT: Pena, Carol E. A
; APPLICANT: Futak, Katarzyna
; APPLICANT: Groese, William M.
; APPLICANT: Alsobrook II, John P.
; APPLICANT: Lepley, Denise M.
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Burgess, Catherine E.
; TITLE OR INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-258
; CURRENT FILING DATE: US/10/072.012
; PRIOR APPLICATION NUMBER: 2002-01-31
; PRIOR FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 60/265,102
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/265,514
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,412
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,395
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/266,406
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/266,767
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/267,057
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/266,975
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/267,459
; PRIOR FILING DATE: 2001-02-08
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1391
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 419
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-072-012-419
Query Match          99.4%; Score 3760; DB 15; Length 855;
Best Local Similarity 99.4%; Pred. No. 2,2e-275;
Matches 679; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
Db 293 LCGTTPSYNLTTFHSSQNVLLITLITWERRHHPGFATFPQLPRMSSCGRLRAQGTFN 352
Qy 181 SPYPGHYHPNIDCTNMIIEVNNQHVVRKFFYLLEPRRACGTCPDYVEINGEKYCGE 240
Db 353 SPYPGHYHPNIDCTNMIIEVNNQHVVRKFFYLLEPRRACGTCPDYVEINGEKYCGE 412
Qy 241 RSQFVVTNSNKTITVRHSDQSYTDGFLAEVLSYSSDPCPOQFTCRTRCIRKELRCD 300
Db 413 RSQFVVTNSNKTITVRHSDQSYTDGFLAEVLSYSSDPCPOQFTCRTRCIRKELRCD 472
Qy 301 GMADCTDHSDELNCSGAGHFTCKNFKCPLFWVCDSDVNDGDNDEQSCSPAQTRFC 360
Db 473 GMADCTDHSDELNCSGAGHFTCKNFKCPLFWVCDSDVNDGDNDEQSCSPAQTRFC 532
Qy 361 SNGKCLSKSQQCKGKDCGSGDEASCPKNNVYTCRTHRCNLGCLSGNRECKED 420
Db 533 SNGKCLSKSQQCKGKDCGSGDEASCPKNNVYTCRTHRCNLGCLSGNRECKED 592
Qy 421 CSDGDEKDCDCGLRSTTRQARVVGTDADGEMPMQVSLHALGQGHICGASLISPMVLV 480
Db 593 CSDGDEKDCDCGLRSTTRQARVVGTDADGEMPMQVSLHALGQGHICGASLISPMVLV 652
Qy 481 SAACVYIDRGFRYSDFPTQWTAFLGHDQSORSAPOVQERRLKRIISHPFNDFTPYDI 540
Db 653 SAACVYIDRGFRYSDFPTQWTAFLGHDQSORSAPOVQERRLKRIISHPFNDFTPYDI 712
Qy 541 ALLELEKPAEYSSMWRTICLPDASHVPPAKAIIWYMGHTQYGGTALILQGEIRVIN 600
Db 713 ALLELEKPAEYSSMWRTICLPDASHVPPAKAIIWYMGHTQYGGTALILQGEIRVIN 772
Qy 601 QTTCENTLPPQITPRMNCVFLSGVDSGCGSDGSPLSVEADGRIFQAGVSMGDSCAQ 660
Db 773 QTTCENTLPPQITPRMNCVFLSGVDSGCGSDGSPLSVEADGRIFQAGVSMGDSCAQ 832
Qy 661 RNKRGVYTRLPFRDWIKENTGV 683
Db 833 RNKRGVYTRLPFRDWIKENTGV 855

RESULT 6
US-10-072-012-354
; Sequence 354, Application US/10072012
; Publication No. US20040033493A1
; GENERAL INFORMATION:
; APPLICANT: Tchernev, Velizar
; APPLICANT: Spytek, Kimberly
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Patuturajan, Meera
; APPLICANT: Shimkets, Richard
; APPLICANT: Li, Li
; APPLICANT: Gangoli, Beha
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Anderson, David W.
; APPLICANT: Rastelli, Luca
; APPLICANT: Miller, Charles E.
; APPLICANT: Gerlach, Valerie E.
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Colman, Steven D.
; APPLICANT: Wolenc, Adam R.
; APPLICANT: Pena, Carol E. A
; APPLICANT: Futak, Katarzyna
; APPLICANT: Groese, William M.
; APPLICANT: Alsobrook II, John P.
; APPLICANT: Lepley, Denise M.
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Burgess, Catherine E.
; TITLE OR INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-258
; CURRENT FILING DATE: US/10/072.012
; PRIOR APPLICATION NUMBER: 2002-01-31
; PRIOR FILING DATE: 2001-01-30
```

;; PRIOR APPLICATION NUMBER: 60/265,514
;; PRIOR FILING DATE: 2001-01-31
;; PRIOR APPLICATION NUMBER: 60/265,517
;; PRIOR FILING DATE: 2001-01-31
;; PRIOR APPLICATION NUMBER: 60/265,412
;; PRIOR FILING DATE: 2001-01-31
;; PRIOR APPLICATION NUMBER: 60/265,395
;; PRIOR FILING DATE: 2001-01-31
;; PRIOR APPLICATION NUMBER: 60/266,406
;; PRIOR FILING DATE: 2001-02-02
;; PRIOR APPLICATION NUMBER: 60/266,767
;; PRIOR FILING DATE: 2001-02-05
;; PRIOR APPLICATION NUMBER: 60/267,057
;; PRIOR FILING DATE: 2001-02-07
;; PRIOR APPLICATION NUMBER: 60/266,975
;; PRIOR FILING DATE: 2001-02-07
;; PRIOR APPLICATION NUMBER: 60/267,459
;; PRIOR FILING DATE: 2001-02-08
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 1391
;; SOFTWARE: Patentln Ver. 2.1
;; SEQ ID NO 354
;; LENGTH: 855
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-072-012-354

Query Match 99.1%; Score 3756; DB 15; Length 855;

Best Local Similarity 99.3%; Pred. No. 4.5e-275; Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAEERVVMLPPRARSLKSFVVTSVVAPPTDSKTVOQTQDNSCSFGHLHARGVELMRFTTPG 60
DB 173 MAEERVVMLPPRARSLKSFVVTSVVAPPTDSKTVOQTQDNSCSFGHLHARGVELMRFTTPG 232
QY 61 FPPSPYPAHARCQWALRGDADSVLSLTFRSPDLASCBERSDLYVTYNTLSPMEPALVQ 120
DB 233 FPPSPYPAHARCQWALRGDADSVLSLTFRSPDLASCBERSDLYVTYNTLSPMEPALVQ 292
QY 121 LCGTTPSYVLTFFHSSQNVLLTLITNTERRHPQFEATFQLPMSCGRLRAQSTFN 180
DB 293 LCGTTPSYVLTFFHSSQNVLLTLITNTERRHPQFEATFQLPMSCGRLRAQSTFN 352
QY 181 SPYYPGHYPNIDCTWNIIEVNNQHVVRPKFYLLEPRACGTCPRDYVEINGEKYCGE 240
DB 353 SPYYPGHYPNIDCTWNIIEVNNQHVVRPKFYLLEPRACGTCPRDYVEINGEKYCGE 412
QY 241 RSQFVVTSSNKTIVRHSQSYTDTGFLAEVLSYDSDPCQGFCTRGCRIRKELRCD 300
DB 413 RSQFVVTSSNKTIVRHSQSYTDTGFLAEVLSYDSDPCQGFCTRGCRIRKELRCD 472
QY 301 GMAADCTHSDLELNCSCDAGHOFCKNFKCKPLFWVCDVNDGNSDEGSCCPAOTFRC 360
DB 473 GMAADCTHSDLELNCSCDAGHOFCKNFKCKPLFWVCDVNDGNSDEGSCCPAOTFRC 532
QY 361 SNGKCLSKSQQCKNGKDCGDSDEASCPKXNVVYCTKHTYRCNLGLCLSGNPECCKED 420
DB 533 SNGKCLSKSQQCKNGKDCGDSDEASCPKXNVVYCTKHTYRCNLGLCLSGNPECCKED 592
QY 421 CSNGSDEKDDCGIRSTRROARVVGTDADGEPWVSLHALGQGHICGASLSPWLV 480
DB 593 CSNGSDEKDDCGIRSTRROARVVGTDADGEPWVSLHALGQGHICGASLSPWLV 652
QY 481 SAHACYIDDRGFRTDPTQWTAFLGLHDOSORSAFQOEKRLKRIISHPFNDFTPDYDI 540
DB 653 SAHACYIDDRGFRTDPTQWTAFLGLHDOSORSAFQOEKRLKRIISHPFNDFTPDYDI 712
QY 541 ALLELEKPAEYSSMWREICLPDASHVPAGKAIWVTGHTOYGGTGALLIQRKEIRVIN 600
DB 713 ALLELEKPAEYSSMWREICLPDASHVPAGKAIWVTGHTOYGGTGALLIQRKEIRVIN 772
QY 601 QTTCCENLLPQOITPRMMCVGFLSGVDSCQDSGGPLSSVEADGRITQAGVWSGDCGCAQ 660

DB 773 QTTCCENLLPQOITPRMMCVGFLSGVDSCQDSGGPLSSVEADGRITQAGVWSGDCGCAQ 832
QY 661 RNRKGYVTRLPFLPRDWIKENTGV 683
DB 833 RNRKGYVTRLPFLPRDWIKENTGV 855

RESULT 7

US-10-072-012-420
;; Sequence 420, Application US/10072012
;; Publication No. US2004003493A1
;; GENERAL INFORMATION:
;; APPLICANT: Tchernev, Velizar
;; APPLICANT: Spytek, Kimberly
;; APPLICANT: Zerhusen, Bryan
;; APPLICANT: Patursajan, Meera
;; APPLICANT: Shinkete, Richard
;; APPLICANT: Li, Li
;; APPLICANT: Gangolli, Beba
;; APPLICANT: Padigaru, Muralidhara
;; APPLICANT: Anderson, David W.
;; APPLICANT: Rastelli, Luca
;; APPLICANT: Miller, Charles E.
;; APPLICANT: Gerlach, Valerie
;; APPLICANT: Taupier Jr, Raymond J.
;; APPLICANT: Gusev, Vladimir Y.
;; APPLICANT: Colman, Steven D.
;; APPLICANT: Wolenc, Adam R.
;; APPLICANT: Pena, Carol E. A.
;; APPLICANT: Furtak, Katarzyna
;; APPLICANT: Grosse, William M.
;; APPLICANT: Alsobrook II, John P.
;; APPLICANT: Lepley, Denise M.
;; APPLICANT: Rieger, Daniel K.
;; APPLICANT: Burgess, Catherine E.
;; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
;; FILE REFERENCE: 21402-258
;; CURRENT APPLICATION NUMBER: US/10/072,012
;; CURRENT FILING DATE: 2002-01-31
;; PRIOR APPLICATION NUMBER: 60/265,102
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: 60/265,514
;; PRIOR FILING DATE: 2001-01-31
;; PRIOR APPLICATION NUMBER: 60/265,517
;; PRIOR FILING DATE: 2001-01-31
;; PRIOR APPLICATION NUMBER: 60/265,412
;; PRIOR FILING DATE: 2001-01-31
;; PRIOR APPLICATION NUMBER: 60/265,395
;; PRIOR FILING DATE: 2001-01-31
;; PRIOR APPLICATION NUMBER: 60/266,406
;; PRIOR FILING DATE: 2001-02-02
;; PRIOR APPLICATION NUMBER: 60/266,767
;; PRIOR FILING DATE: 2001-02-05
;; PRIOR APPLICATION NUMBER: 60/267,057
;; PRIOR FILING DATE: 2001-02-07
;; PRIOR APPLICATION NUMBER: 60/266,975
;; PRIOR FILING DATE: 2001-02-07
;; PRIOR APPLICATION NUMBER: 60/267,459
;; PRIOR FILING DATE: 2001-02-08
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 1391
;; SOFTWARE: Patentln Ver. 2.1
;; SEQ ID NO 420
;; LENGTH: 855
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-072-012-420

Query Match 99.1%; Score 3756; DB 15; Length 855;

Best Local Similarity 99.3%; Pred. No. 4.5e-275; Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAEERVVMLPPRARSLKSFVVTSVVAPPTDSKTVOQTQDNSCSFGHLHARGVELMRFTTPG 60

```

Db 173 MAERVVMLPPRARSLSFVVTSVVAFPTDSKTQRTQDNCSFGHARGVELMRFTTGG 232
Qy 61 PPDSPPYPAHARCCQALRGDADSVLSTFRSFDLASCDERSGDLVTVNTTSPMEPHALVQ 120
Db 233 PPDSPPYPAHARCCQALRGDADSVLSTFRSFDLASCDERSGDLVTVNTTSPMEPHALVQ 292
Qy 121 LCGTTPPSYNLTFFHSSQNVLLITLITNTERHHPGEATFQLPMSSCGGRLRAQGTEN 180
Db 293 LCGTTPPSYNLTFFHSSQNVLLITLITNTERHHPGEATFQLPMSSCGGRLRAQGTEN 352
Qy 181 SPYYPGHYPNIDCTWNIIEVNNQHVAKRFKFFYLEPRRACGTCPCXDYVEINKEKXCGE 240
Db 353 SPYYPGHYPNIDCTWNIIEVNNQHVAKRFKFFYLEPRRACGTCPCXDYVEINKEKXCGE 412
Qy 241 RSQFVTVTSNKKITVRPHSDQSYTDGTGFLAEYLSYSSDPCPGQFTCRTRGCIKRELCD 300
Db 413 RSQFVTVTSNKKITVRPHSDQSYTDGTGFLAEYLSYSSDPCPGQFTCRTRGCIKRELCD 472
Qy 301 GMADCTHSDDELNCSGAGHOFCTCKNFKCPLFWVCDSVNDCCGNSDEQSCSPAQTFRG 360
Db 473 GMADCTHSDDELNCSGAGHOFCTCKNFKCPLFWVCDSVNDCCGNSDEQSCSPAQTFRG 532
Qy 361 SNGKCLSKSQOQCNKDCGSDGDEASCPKVVVYCTGHTYRCLNGLCLSKGNPECDGKED 420
Db 533 SNGKCLSKSQOQCNKDCGSDGDEASCPKVVVYCTGHTYRCLNGLCLSKGNPECDGKED 592
Qy 421 CSDGSDERKDCDGLRSTTRQARVVGTDADGEMPMQVSLHALGQGHICGASLISPMWLV 480
Db 593 CSDGSDERKDCDGLRSTTRQARVVGTDADGEMPMQVSLHALGQGHICGASLISPMWLV 652
Qy 481 SAACICYIDDRGFRSDPTQWTAFLGLHDQSORSAFGVERLRLKIIISHPFNDFTFYDI 540
Db 653 SAACICYIDDRGFRSDPTQWTAFLGLHDQSORSAFGVERLRLKIIISHPFNDFTFYDI 712
Qy 541 ALLELEKPAEYXSWVRPCLPDASHVFPAGKAIWVTGHTYRCLNGLCLSKGNPECDGKED 600
Db 713 ALLELEKPAEYXSWVRPCLPDASHVFPAGKAIWVTGHTYRCLNGLCLSKGNPECDGKED 772
Qy 601 QTTCENTLLPQOITPRMNCVGLSGGVDSQCGSDGGPLSSVEADGRIFQAGVSWMGDCAQ 660
Db 773 QTTCENTLLPQOITPRMNCVGLSGGVDSQCGSDGGPLSSVEADGRIFQAGVSWMGDCAQ 832
Qy 661 RNKRGVYTRRLPLFRDWIKENTGV 683
Db 833 RNKRGVYTRRLPLFRDWIKENTGV 855

```

```

APPLICANT: Shenoy, Suresh G
APPLICANT: Anderson, David W
APPLICANT: Padigaru, Muraidhara
APPLICANT: Taupier Jr, Raymond J
APPLICANT: Miller, Charles E
APPLICANT: Eileen, Andrew J
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
FILE REFERENCE: 21402-235
CURRENT APPLICATION NUMBER: US/10/037,417
PRIOR APPLICATION NUMBER: 2002-09-20
PRIOR APPLICATION NUMBER: 60/260,018
PRIOR FILING DATE: 2001-01-05
PRIOR APPLICATION NUMBER: 60/260,360
PRIOR FILING DATE: 2001-01-08
PRIOR APPLICATION NUMBER: 60/272,411
PRIOR FILING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: 60/272,817
PRIOR FILING DATE: 2001-03-02
PRIOR APPLICATION NUMBER: 60/291,186
PRIOR FILING DATE: 2001-05-15
PRIOR APPLICATION NUMBER: 60/303,231
PRIOR FILING DATE: 2001-07-05
PRIOR APPLICATION NUMBER: 60/305,060
PRIOR FILING DATE: 2001-07-12
PRIOR APPLICATION NUMBER: 60/318,405
PRIOR FILING DATE: 2001-09-10
PRIOR APPLICATION NUMBER: 60/318,700
PRIOR FILING DATE: 2001-09-12
NUMBER OF SEQ ID NOS: 227
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 132
LENGTH: 855
TYPE: PRT
ORGANISM: Homo sapiens
US-10-037-417-132
Query Match 99.1%; Score 3756; DB 15; Length 855;
Best Local Similarity 99.3%; Pred. No. 4.5e-275;
Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Qy 1 MAERVVMLPPRARSLSFVVTSVVAFPTDSKTQRTQDNCSFGHARGVELMRFTTGG 60
Db 173 MAERVVMLPPRARSLSFVVTSVVAFPTDSKTQRTQDNCSFGHARGVELMRFTTGG 232
Qy 61 PPDSPPYPAHARCCQALRGDADSVLSTFRSFDLASCDERSGDLVTVNTTSPMEPHALVQ 120
Db 233 PPDSPPYPAHARCCQALRGDADSVLSTFRSFDLASCDERSGDLVTVNTTSPMEPHALVQ 292
Qy 121 LCGTTPPSYNLTFFHSSQNVLLITLITNTERHHPGEATFQLPMSSCGGRLRAQGTEN 180
Db 293 LCGTTPPSYNLTFFHSSQNVLLITLITNTERHHPGEATFQLPMSSCGGRLRAQGTEN 352
Qy 181 SPYYPGHYPNIDCTWNIIEVNNQHVAKRFKFFYLEPRRACGTCPCXDYVEINKEKXCGE 240
Db 353 SPYYPGHYPNIDCTWNIIEVNNQHVAKRFKFFYLEPRRACGTCPCXDYVEINKEKXCGE 412
Qy 241 RSQFVTVTSNKKITVRPHSDQSYTDGTGFLAEYLSYSSDPCPGQFTCRTRGCIKRELCD 300
Db 413 RSQFVTVTSNKKITVRPHSDQSYTDGTGFLAEYLSYSSDPCPGQFTCRTRGCIKRELCD 472
Qy 301 GMADCTHSDDELNCSGAGHOFCTCKNFKCPLFWVCDSVNDCCGNSDEQSCSPAQTFRG 360
Db 473 GMADCTHSDDELNCSGAGHOFCTCKNFKCPLFWVCDSVNDCCGNSDEQSCSPAQTFRG 532
Qy 361 SNGKCLSKSQOQCNKDCGSDGDEASCPKVVVYCTGHTYRCLNGLCLSKGNPECDGKED 420
Db 533 SNGKCLSKSQOQCNKDCGSDGDEASCPKVVVYCTGHTYRCLNGLCLSKGNPECDGKED 592
Qy 421 CSDGSDERKDCDGLRSTTRQARVVGTDADGEMPMQVSLHALGQGHICGASLISPMWLV 480
Db 593 CSDGSDERKDCDGLRSTTRQARVVGTDADGEMPMQVSLHALGQGHICGASLISPMWLV 652
Qy 481 SAACICYIDDRGFRSDPTQWTAFLGLHDQSORSAFGVERLRLKIIISHPFNDFTFYDI 540

```

```
Db 653 SAACHYIDDDGFRYSPTQMTVFLGLHDQSRSAPGVQERRLKRIISHPFENDFTFDYDI 712
Qy 541 ALLELEKPAEYSSMWPRICLPDASHVPAGKAIWVTGMGHTQYGGTALILQGEIRVIN 600
Db 713 ALLELEKPAEYSSMWPRICLPDASHVPAGKAIWVTGMGHTQYGGTALILQGEIRVIN 772
Qy 601 QTTCEMLLPQOITPRMNCVGFSLSGVDSCQDSCGGLSVYADGRIFQAGVSWGDCAQ 660
Db 773 QTTCEMLLPQOITPRMNCVGFSLSGVDSCQDSCGGLSVYADGRIFQAGVSWGDCAQ 832
Qy 661 RNKRGVYTRLPFRDWIKENTGV 683
Db 833 RNKRGVYTRLPFRDWIKENTGV 855

RESULT 9
US-10-072-012-44
; Sequence 44, Application US/10072012
; Publication No. US2004003493A1
; GENERAL INFORMATION:
; APPLICANT: Tchernev, Velizar
; APPLICANT: Spytek, Kimberly
; APPLICANT: Zehusen, Bryan
; APPLICANT: Patuturajan, Meera
; APPLICANT: Shmukets, Richard
; APPLICANT: Li, Li
; APPLICANT: Gangolli, Esna
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Anderson, David W.
; APPLICANT: Rastelli, Luca
; APPLICANT: Miller, Charles E.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Coleman, Steven D.
; APPLICANT: Wolenc, Adam R.
; APPLICANT: Pena, Carol E. A
; APPLICANT: Putak, Katarzyna
; APPLICANT: Groesse, William M.
; APPLICANT: Alsobrook II, John P.
; APPLICANT: Lepley, Denise M.
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Burgees, Catherine E.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-258
; CURRENT APPLICATION NUMBER: US/10/072,012
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 60/265,102
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/265,514
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,412
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,395
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/266,406
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/266,767
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/267,057
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/266,975
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/267,459
; PRIOR FILING DATE: 2001-02-08
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1391
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 44
; LENGTH: 757
```

```
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-072-012-44
Query Match 99.1%; Score 3754; DB 15; Length 757;
Best Local Similarity 99.3%; Pred. No. 5.5e-275;
Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 MAERVYMLPPRRARSLKSFVVTSVVAFPTDSKTQVORTODNSCSFGHLHARGVEMLRFTTPG 60
Db 75 MAERVYMLPPRRARSLKSFVVTSVVAFPTDSKTQVORTODNSCSFGHLHARGVEMLRFTTPG 134
Qy 61 FPDSPYPAHARCOMALRGDADSVLSTFRSFDIASCDERGSGLVYVNTLSPEMHPALVQ 120
Db 135 FPDSPYPAHARCOMALRGDADSVLSTFRSFDIASCDERGSGLVYVNTLSPEMHPALVQ 194
Qy 121 LCGTTPPSVYLTFRSSONVLLITLINTBRHHPFATFQULPRMSSCGRLRKAGTFN 180
Db 195 LCGTTPPSVYLTFRSSONVLLITLINTBRHHPFATFQULPRMSSCGRLRKAGTFN 254
Qy 181 SPYYPGHYPNIDCTWNIIEVPNNQHVAFKFFYLLEPRRACGTCPKDYVEINGEKYCGE 240
Db 255 SPYYPGHYPNIDCTWNIIEVPNNQHVAFKFFYLLEPGVPATCPKDYVEINGEKYCGE 314
Qy 241 RSQFVVTSSNKKITVFRHSDQSYDTGTFLAEVLSYDSDPCPQOFTCRGCIKRLKCD 300
Db 315 RSQFVVTSSNKKITVFRHSDQSYDTGTFLAEVLSYDSDPCPQOFTCRGCIKRLKCD 374
Qy 301 GMADCTDHSDELNCSGCDAGHOFCKNPKFCKPLFWVCDVNDGSDNSDEQSCGPAQTFRC 360
Db 375 GMADCTDHSDELNCSGCDAGHOFCKNPKFCKPLFWVCDVNDGSDNSDEQSCGPAQTFRC 434
Qy 361 SNGCKLSKSOQCKGKDCGDSDEASCPKVVVCTKHTYRCINGLCLSKGNBECQKED 420
Db 435 SNGCKLSKSOQCKGKDCGDSDEASCPKVVVCTKHTYRCINGLCLSKGNBECQKED 494
Qy 421 CSDGSEKDCDGLRFTQARVVGSTDADEGEPMQVSLHALGQHICGASLISPMWLV 480
Db 495 CSDGSEKDCDGLRFTQARVVGSTDADEGEPMQVSLHALGQHICGASLISPMWLV 554
Qy 481 SAACHYIDDRGFYSPTQMTAFGLHDQSRSAPGVQERRLKRIISHPFENDFTFDYDI 540
Db 555 SAACHYIDDRGFYSPTQMTAFGLHDQSRSAPGVQERRLKRIISHPFENDFTFDYDI 614
Qy 541 ALLELEKPAEYSSMWPRICLPDASHVPAGKAIWVTGMGHTQYGGTALILQGEIRVIN 600
Db 615 ALLELEKPAEYSSMWPRICLPDASHVPAGKAIWVTGMGHTQYGGTALILQGEIRVIN 674
Qy 601 QTTCEMLLPQOITPRMNCVGFSLSGVDSCQDSCGGLSVYADGRIFQAGVSWGDCAQ 660
Db 675 QTTCEMLLPQOITPRMNCVGFSLSGVDSCQDSCGGLSVYADGRIFQAGVSWGDCAQ 734
Qy 661 RNKRGVYTRLPFRDWIKENTGV 683
Db 735 RNKRGVYTRLPFRDWIKENTGV 757

RESULT 10
US-09-776-191-2
; Sequence 2, Application US/09776191
; Publication No. US20030119168A1
; GENERAL INFORMATION:
; APPLICANT: Edwin L. Madison
; APPLICANT: Edgar O. Ong
; APPLICANT: Jiumn-Chern Yeh
; APPLICANT: Corvas International, Inc.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; TITLE OF INVENTION: TRANSMEMBRANE SERINE PROTEASES, THE ENCODED PROTEINS AND
; FILE REFERENCE: 24745-1607
; CURRENT APPLICATION NUMBER: US/09/776,191
; CURRENT FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/213,124
```

PRIOR FILING DATE: 2000-06-22
PRIOR APPLICATION NUMBER: 60/234,840
PRIOR FILING DATE: 2000-06-22
PRIOR APPLICATION NUMBER: 60/179,982
PRIOR FILING DATE: 2000-02-03
PRIOR APPLICATION NUMBER: 60/183,542
PRIOR FILING DATE: 2000-02-18
PRIOR APPLICATION NUMBER: 09/657,968
PRIOR FILING DATE: 2000-02-08
NUMBER OF SEQ ID NOS: 72
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 855
TYPE: PRT
ORGANISM: Homo Sapien
US-09-776-191-2

Query Match 99.1%; Score 3754; DB 10; Length 855;
Best Local Similarity 99.3%; Pred. No. 6.4e-275;
Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAERVYMLPPRARSLSFVVTGVAFPDSTKVQRTQDNCSCFGLHARGVELMRFTTPG 60
DB 173 MAERVYMLPPRARSLSFVVTGVAFPDSTKVQRTQDNCSCFGLHARGVELMRFTTPG 232
QY 61 PPDSPPAHARCCQALRGDADSVLSTFRSFDLASCDERGSIDLVTYNTLSPMEPHALVQ 120
DB 233 PPDSPPAHARCCQALRGDADSVLSTFRSFDLASCDERGSIDLVTYNTLSPMEPHALVQ 292
QY 121 LCGTTPSYNLTFHSSQNVLLITLITNTERRHGFEATFQLPMSGCCGLRKAQGTEN 180
DB 293 LCGTTPSYNLTFHSSQNVLLITLITNTERRHGFEATFQLPMSGCCGLRKAQGTEN 352
QY 181 SPYYPGHYPNIDCTWNIIEVNNQHVKRFKFLLEBRACGTCPKDYVEINGEKYCGE 240
DB 353 SPYYPGHYPNIDCTWNIIEVNNQHVKRFKFLLEBRACGTCPKDYVEINGEKYCGE 412
QY 241 RSQGVVTSNKNKIVTRPHSDQSYTDGTGLAEVLSYDSDPCPGQFTCTGRCKIRKELRCD 300
DB 413 RSQGVVTSNKNKIVTRPHSDQSYTDGTGLAEVLSYDSDPCPGQFTCTGRCKIRKELRCD 472
QY 301 GMADCTHSDLELNCSCDAGHOFCKNFKCKPLFWVCDSDVNDGNSDEQSCCPAQTFRC 360
DB 473 GMADCTHSDLELNCSCDAGHOFCKNFKCKPLFWVCDSDVNDGNSDEQSCCPAQTFRC 532
QY 361 SNGKLSKSQOCNGKDCGDSDEASCPKXNVVTCIKHYRCLNGLCLSKGNPECDEKED 420
DB 533 SNGKLSKSQOCNGKDCGDSDEASCPKXNVVTCIKHYRCLNGLCLSKGNPECDEKED 592
QY 421 CSDGSDERKDCDGLRSFTROARVVGTTADGEMPMQVSLHALQGHICGASLISPMVLV 480
DB 593 CSDGSDERKDCDGLRSFTROARVVGTTADGEMPMQVSLHALQGHICGASLISPMVLV 652
QY 481 SAACHYIDDDGFRYSPTQMTAFGLHDQSORAPGVQERLRKRIISHPFNDFTFYDI 540
DB 653 SAACHYIDDDGFRYSPTQMTAFGLHDQSORAPGVQERLRKRIISHPFNDFTFYDI 712
QY 541 ALLELEKPAEYSSNWRPCLPDASHVPAGKAIWVTMGHTQYGGTALILQKEIRVIN 600
DB 713 ALLELEKPAEYSSNWRPCLPDASHVPAGKAIWVTMGHTQYGGTALILQKEIRVIN 772
QY 601 QTTCEMLLPQOITPRMNCVFLSGGVDSCGDSCGGLPSVEADGRIFQAGVSWGDCAQ 660
DB 773 QTTCEMLLPQOITPRMNCVFLSGGVDSCGDSCGGLPSVEADGRIFQAGVSWGDCAQ 832
QY 661 RNKPGVYTRLPFRDWIKENTGV 683
DB 833 RNKPGVYTRLPFRDWIKENTGV 855

RESULT 11
US-10-099-700A-2
; Sequence 2, Application US/10099700A

Publication No. US20030008372A1
GENERAL INFORMATION:
APPLICANT: Edwin L. Madison
APPLICANT: Edgar O. Ong
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING TRANSMEMBRANE SERINE PROTEASE 7,
FILE REFERENCE: 24745-1613
CURRENT APPLICATION NUMBER: US/10/099,700A
CURRENT FILING DATE: 2002-05-24
PRIOR APPLICATION NUMBER: 60/275,592
PRIOR FILING DATE: 2001-03-13
NUMBER OF SEQ ID NOS: 22
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 855
TYPE: PRT
ORGANISM: Homo Sapien
US-10-099-700A-2

Query Match 99.1%; Score 3754; DB 14; Length 855;
Best Local Similarity 99.3%; Pred. No. 6.4e-275;
Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAERVYMLPPRARSLSFVVTGVAFPDSTKVQRTQDNCSCFGLHARGVELMRFTTPG 60
DB 173 MAERVYMLPPRARSLSFVVTGVAFPDSTKVQRTQDNCSCFGLHARGVELMRFTTPG 232
QY 61 PPDSPPAHARCCQALRGDADSVLSTFRSFDLASCDERGSIDLVTYNTLSPMEPHALVQ 120
DB 233 PPDSPPAHARCCQALRGDADSVLSTFRSFDLASCDERGSIDLVTYNTLSPMEPHALVQ 292
QY 121 LCGTTPSYNLTFHSSQNVLLITLITNTERRHGFEATFQLPMSGCCGLRKAQGTEN 180
DB 293 LCGTTPSYNLTFHSSQNVLLITLITNTERRHGFEATFQLPMSGCCGLRKAQGTEN 352
QY 181 SPYYPGHYPNIDCTWNIIEVNNQHVKRFKFLLEBRACGTCPKDYVEINGEKYCGE 240
DB 353 SPYYPGHYPNIDCTWNIIEVNNQHVKRFKFLLEBRACGTCPKDYVEINGEKYCGE 412
QY 241 RSQGVVTSNKNKIVTRPHSDQSYTDGTGLAEVLSYDSDPCPGQFTCTGRCKIRKELRCD 300
DB 413 RSQGVVTSNKNKIVTRPHSDQSYTDGTGLAEVLSYDSDPCPGQFTCTGRCKIRKELRCD 472
QY 301 GMADCTHSDLELNCSCDAGHOFCKNFKCKPLFWVCDSDVNDGNSDEQSCCPAQTFRC 360
DB 473 GMADCTHSDLELNCSCDAGHOFCKNFKCKPLFWVCDSDVNDGNSDEQSCCPAQTFRC 532
QY 361 SNGKLSKSQOCNGKDCGDSDEASCPKXNVVTCIKHYRCLNGLCLSKGNPECDEKED 420
DB 533 SNGKLSKSQOCNGKDCGDSDEASCPKXNVVTCIKHYRCLNGLCLSKGNPECDEKED 592
QY 421 CSDGSDERKDCDGLRSFTROARVVGTTADGEMPMQVSLHALQGHICGASLISPMVLV 480
DB 593 CSDGSDERKDCDGLRSFTROARVVGTTADGEMPMQVSLHALQGHICGASLISPMVLV 652
QY 481 SAACHYIDDDGFRYSPTQMTAFGLHDQSORAPGVQERLRKRIISHPFNDFTFYDI 540
DB 653 SAACHYIDDDGFRYSPTQMTAFGLHDQSORAPGVQERLRKRIISHPFNDFTFYDI 712
QY 541 ALLELEKPAEYSSNWRPCLPDASHVPAGKAIWVTMGHTQYGGTALILQKEIRVIN 600
DB 713 ALLELEKPAEYSSNWRPCLPDASHVPAGKAIWVTMGHTQYGGTALILQKEIRVIN 772
QY 601 QTTCEMLLPQOITPRMNCVFLSGGVDSCGDSCGGLPSVEADGRIFQAGVSWGDCAQ 660
DB 773 QTTCEMLLPQOITPRMNCVFLSGGVDSCGDSCGGLPSVEADGRIFQAGVSWGDCAQ 832
QY 661 RNKPGVYTRLPFRDWIKENTGV 683
DB 833 RNKPGVYTRLPFRDWIKENTGV 855

RESULT 12

US-10-190-030B-2
; Sequence 2, Application US/10190030B
; Publication No. US20030134298A1
; GENERAL INFORMATION:
; APPLICANT: Madison, Edwin
; APPLICANT: Ong, Edgar
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING A
; TITLE OF INVENTION: TRANSMEMBRANE SERINE PROTEASE 20, THE ENCODED POLYPEPTIDES AND
; FILE REFERENCE: 24745-1618
; CURRENT APPLICATION NUMBER: US/10/190,030B
; CURRENT FILING DATE: 2002-07-03
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PR1
; ORGANISM: Homo Sapien
US-10-190-030B-2

Query Match 99.1%; Score 3754; DB 14; Length 855;

Best Local Similarity 99.3%; Pred. No. 6.4e-275;
Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAERVVMLPPRARSLSKSPVVTSVVAFPTDSKTVOQTODNSCSFGLHARGVELMRFTTGG 60
DB 173 MAERVVMLPPRARSLSKSPVVTSVVAFPTDSKTVOQTODNSCSFGLHARGVELMRFTTGG 232
QY 61 PPDSPPYPAHARCOMALRGDADSVLSTFRSFDLASCDERGSIDLVTYNTLSPMEPHALVQ 120
DB 233 PPDSPPYPAHARCOMALRGDADSVLSTFRSFDLASCDERGSIDLVTYNTLSPMEPHALVQ 292
QY 121 LCGTTPSYNLTFHSSONVLLITLITNTERRHGFEATFPOLPRMSSCGRLKAQGTEN 180
DB 293 LCGTTPSYNLTFHSSONVLLITLITNTERRHGFEATFPOLPRMSSCGRLKAQGTEN 352
QY 181 SPYPGHYPPNIDCTNIEVPNNQHVKRFKFFYLLPBRACGTCRKYVEINKEKXCGE 240
DB 353 SPYPGHYPPNIDCTNIEVPNNQHVKRFKFFYLLPBRACGTCRKYVEINKEKXCGE 412
QY 241 RSQFVVTNSNKTITVRFHSDQSTYDTGFLAEYLSYDSDDPCPGQFCTRGRCIRKELRCD 300
DB 413 RSQFVVTNSNKTITVRFHSDQSTYDTGFLAEYLSYDSDDPCPGQFCTRGRCIRKELRCD 472
QY 301 GMAADCTHSDDELNCSCDAGHOFCKNFKCKPLFWVCDVNDGDNDEQSCCPAQTFRC 360
DB 473 GMAADCTHSDDELNCSCDAGHOFCKNFKCKPLFWVCDVNDGDNDEQSCCPAQTFRC 532
QY 361 SNGKCLSKSQOQCNKGKDCGDSDEASCPRKNVVTCTKHTYRCINGLCLSKGNPECDGKED 420
DB 533 SNGKCLSKSQOQCNKGKDCGDSDEASCPRKNVVTCTKHTYRCINGLCLSKGNPECDGKED 592
QY 421 CSQDSBCKDCDGLRSTFRARVVGSTDADEGMPQVSIHALGOGHICGASLISPMWLV 480
DB 593 CSQDSBCKDCDGLRSTFRARVVGSTDADEGMPQVSIHALGOGHICGASLISPMWLV 652
QY 481 SAHACYIDRGFRYSPTQWTAFLGLHDOSQSAAPGQERRLKRIISHPFNDFTFDYDI 540
DB 653 SAHACYIDRGFRYSPTQWTAFLGLHDOSQSAAPGQERRLKRIISHPFNDFTFDYDI 712
QY 541 ALLELEKPAEYSSMWRPCLPDAASHVPAGKAIWVTGWHGTQYGGTALILQKEIRVIN 600
DB 713 ALLELEKPAEYSSMWRPCLPDAASHVPAGKAIWVTGWHGTQYGGTALILQKEIRVIN 772
QY 601 QTTCEMLLPQOITPRMMCVGFLSGVDSGCGDSGGLSVYADGRIFQAGVSWGDCAQ 660
DB 773 QTTCEMLLPQOITPRMMCVGFLSGVDSGCGDSGGLSVYADGRIFQAGVSWGDCAQ 832
QY 661 RNKPGVYTRLLPLFRDWIKENTGV 683
DB 833 RNKPGVYTRLLPLFRDWIKENTGV 855

RESULT 13
US-10-302-840A-2
; Sequence 2, Application US/10302840A
; Publication No. US20030134794A1
; GENERAL INFORMATION:
; APPLICANT: Madison, Edwin L.
; APPLICANT: Ong, Edgar O.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING SERINE PROTEASE CVSP17, THE ENCODED
; TITLE OF INVENTION: POLYPEPTIDES AND METHODS BASED THEREON
; FILE REFERENCE: 24745-1622
; CURRENT APPLICATION NUMBER: US/10/302,840A
; CURRENT FILING DATE: 2003-01-24
; PRIOR APPLICATION NUMBER: 60/332,015
; PRIOR FILING DATE: 2001-11-20
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PR1
; ORGANISM: Homo Sapien
US-10-302-840A-2

Query Match 99.1%; Score 3754; DB 14; Length 855;

Best Local Similarity 99.3%; Pred. No. 6.4e-275;
Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAERVVMLPPRARSLSKSPVVTSVVAFPTDSKTVOQTODNSCSFGLHARGVELMRFTTGG 60
DB 173 MAERVVMLPPRARSLSKSPVVTSVVAFPTDSKTVOQTODNSCSFGLHARGVELMRFTTGG 232
QY 61 PPDSPPYPAHARCOMALRGDADSVLSTFRSFDLASCDERGSIDLVTYNTLSPMEPHALVQ 120
DB 233 PPDSPPYPAHARCOMALRGDADSVLSTFRSFDLASCDERGSIDLVTYNTLSPMEPHALVQ 292
QY 121 LCGTTPSYNLTFHSSONVLLITLITNTERRHGFEATFPOLPRMSSCGRLKAQGTEN 180
DB 293 LCGTTPSYNLTFHSSONVLLITLITNTERRHGFEATFPOLPRMSSCGRLKAQGTEN 352
QY 181 SPYPGHYPPNIDCTNIEVPNNQHVKRFKFFYLLPBRACGTCRKYVEINKEKXCGE 240
DB 353 SPYPGHYPPNIDCTNIEVPNNQHVKRFKFFYLLPBRACGTCRKYVEINKEKXCGE 412
QY 241 RSQFVVTNSNKTITVRFHSDQSTYDTGFLAEYLSYDSDDPCPGQFCTRGRCIRKELRCD 300
DB 413 RSQFVVTNSNKTITVRFHSDQSTYDTGFLAEYLSYDSDDPCPGQFCTRGRCIRKELRCD 472
QY 301 GMAADCTHSDDELNCSCDAGHOFCKNFKCKPLFWVCDVNDGDNDEQSCCPAQTFRC 360
DB 473 GMAADCTHSDDELNCSCDAGHOFCKNFKCKPLFWVCDVNDGDNDEQSCCPAQTFRC 532
QY 361 SNGKCLSKSQOQCNKGKDCGDSDEASCPRKNVVTCTKHTYRCINGLCLSKGNPECDGKED 420
DB 533 SNGKCLSKSQOQCNKGKDCGDSDEASCPRKNVVTCTKHTYRCINGLCLSKGNPECDGKED 592
QY 421 CSQDSBCKDCDGLRSTFRARVVGSTDADEGMPQVSIHALGOGHICGASLISPMWLV 480
DB 593 CSQDSBCKDCDGLRSTFRARVVGSTDADEGMPQVSIHALGOGHICGASLISPMWLV 652
QY 481 SAHACYIDRGFRYSPTQWTAFLGLHDOSQSAAPGQERRLKRIISHPFNDFTFDYDI 540
DB 653 SAHACYIDRGFRYSPTQWTAFLGLHDOSQSAAPGQERRLKRIISHPFNDFTFDYDI 712
QY 541 ALLELEKPAEYSSMWRPCLPDAASHVPAGKAIWVTGWHGTQYGGTALILQKEIRVIN 600
DB 713 ALLELEKPAEYSSMWRPCLPDAASHVPAGKAIWVTGWHGTQYGGTALILQKEIRVIN 772
QY 601 QTTCEMLLPQOITPRMMCVGFLSGVDSGCGDSGGLSVYADGRIFQAGVSWGDCAQ 660
DB 773 QTTCEMLLPQOITPRMMCVGFLSGVDSGCGDSGGLSVYADGRIFQAGVSWGDCAQ 832
QY 661 RNKPGVYTRLLPLFRDWIKENTGV 683
DB 833 RNKPGVYTRLLPLFRDWIKENTGV 855

RESULT 14
US-10-267-219-2
; Sequence 2, Application US/10267219
; Publication No. US20030143219A1
; GENERAL INFORMATION:
; APPLICANT: Madison, Edwin
; APPLICANT: Yeh, Jium-Chern
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING A TRANSMEMBRANE SERINE PROTEASE 2
; TITLE OF INVENTION: ENCODED POLYPEPTIDES AND METHODS BASED THEREON
; FILE REFERENCE: 24745-1621
; CURRENT APPLICATION NUMBER: US/10/267,219
; CURRENT FILING DATE: 2002-10-08
; PRIOR APPLICATION NUMBER: 60/328,530
; PRIOR FILING DATE: 09-Oct-2001
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-267-219-2

Query Match 99.1%; Score 3754; DB 14; Length 855;
Best Local Similarity 99.3%; Pred. No. 6,4e-275;
Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAERVVMLEPPRASLSKSFVVTSVVAFPTDSKTVQRTQDNCSCFGLHARGVELMRFTTPG 60
DB 173 MAERVVMLEPPRASLSKSFVVTSVVAFPTDSKTVQRTQDNCSCFGLHARGVELMRFTTPG 232

QY 61 FPDSPYPAHARQWALRGDADSVSLTFRSFDLASCDERSGDLVTVNTLSPMEPHALVQ 120
DB 233 FPDSPYPAHARQWALRGDADSVSLTFRSFDLASCDERSGDLVTVNTLSPMEPHALVQ 292

QY 121 LCGTYPSPSYNLTFHSSQNVLLITLITNTERRHGFEATFQLPKMSCGRLRKAQGTEN 180
DB 293 LCGTYPSPSYNLTFHSSQNVLLITLITNTERRHGFEATFQLPKMSCGRLRKAQGTEN 352

QY 181 SPYYPGHYPNIDCTWNIIEVNNQHVVRKFKFYLLERPRACGTCPKDYVEINGEKYGE 240
DB 353 SPYYPGHYPNIDCTWNIIEVNNQHVVRKFKFYLLERPGVAGTCPKDYVEINGEKYGE 412

QY 241 RSQFVVTNSNKTIVRFHSDSYTDITGFLAEYLSYDSSDPCPGQFTCRGCIKRLRCD 300
DB 413 RSQFVVTNSNKTIVRFHSDSYTDITGFLAEYLSYDSSDPCPGQFTCRGCIKRLRCD 472

QY 301 GMAUCTHSDDELNCSCDAGHGFCTCKNFKCKPLFWVCDSDVNDGDNDEGSCSPAQTFR 360
DB 473 GMAUCTHSDDELNCSCDAGHGFCTCKNFKCKPLFWVCDSDVNDGDNDEGSCSPAQTFR 532

QY 361 SNGKCLSKSQOQNGKDCGSDGDEASCPKXNVVVTCTKTYRCLNGLCLSKGNPECDKED 420
DB 533 SNGKCLSKSQOQNGKDCGSDGDEASCPKXNVVVTCTKTYRCLNGLCLSKGNPECDKED 592

QY 421 CSDDSDDEKDCCGRLRSFTROARVVGCTADDEGEMPMQVSLHALGQGHICGASLISPMVLV 480
DB 593 CSDDSDDEKDCCGRLRSFTROARVVGCTADDEGEMPMQVSLHALGQGHICGASLISPMVLV 652

QY 481 SAACHYIDDRGFRYSDFPTQMTAFGLHDQSORAPGVQERRLKRIISHPFNDFTFDYDI 540
DB 653 SAACHYIDDRGFRYSDFPTQMTAFGLHDQSORAPGVQERRLKRIISHPFNDFTFDYDI 712

QY 541 ALLELEKPAEYSSWVRPICLPDASHVFPAGKAIWVTGMGHTQYGGTGAALLQKGEIRVIN 600
DB 713 ALLELEKPAEYSSWVRPICLPDASHVFPAGKAIWVTGMGHTQYGGTGAALLQKGEIRVIN 772

QY 601 QTTCEMLLPQOITPRMNCVGLSGGVNSCGDSSGGLSVADGRIFPAGVAVSWGDCAQ 660
DB 773 QTTCEMLLPQOITPRMNCVGLSGGVNSCGDSSGGLSVADGRIFPAGVAVSWGDCAQ 832

QY 661 RNKPGVYTRLPFLFRDWIKENTGV 683

DB 833 RNKPGVYTRLPFLFRDWIKENTGV 855

RESULT 15
US-10-112-221A-2
; Sequence 2, Application US/10112221A
; Publication No. US20030166851A1
; GENERAL INFORMATION:
; APPLICANT: Madison, Edwin
; APPLICANT: Ong, Edgar O.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING A TRANSMEMBRANE SERINE PROTEASE 9,
; TITLE OF INVENTION: ENCODED POLYPEPTIDES AND METHODS BASED THEREON
; FILE REFERENCE: 24745-1615
; CURRENT APPLICATION NUMBER: US/10/112,221A
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: 60/279,228
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: 60/291,501
; PRIOR FILING DATE: 2001-05-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-112-221A-2

Query Match 99.1%; Score 3754; DB 14; Length 855;
Best Local Similarity 99.3%; Pred. No. 6,4e-275;
Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAERVVMLEPPRASLSKSFVVTSVVAFPTDSKTVQRTQDNCSCFGLHARGVELMRFTTPG 60
DB 173 MAERVVMLEPPRASLSKSFVVTSVVAFPTDSKTVQRTQDNCSCFGLHARGVELMRFTTPG 232

QY 61 FPDSPYPAHARQWALRGDADSVSLTFRSFDLASCDERSGDLVTVNTLSPMEPHALVQ 120
DB 233 FPDSPYPAHARQWALRGDADSVSLTFRSFDLASCDERSGDLVTVNTLSPMEPHALVQ 292

QY 121 LCGTYPSPSYNLTFHSSQNVLLITLITNTERRHGFEATFQLPKMSCGRLRKAQGTEN 180
DB 293 LCGTYPSPSYNLTFHSSQNVLLITLITNTERRHGFEATFQLPKMSCGRLRKAQGTEN 352

QY 181 SPYYPGHYPNIDCTWNIIEVNNQHVVRKFKFYLLERPRACGTCPKDYVEINGEKYGE 240
DB 353 SPYYPGHYPNIDCTWNIIEVNNQHVVRKFKFYLLERPGVAGTCPKDYVEINGEKYGE 412

QY 241 RSQFVVTNSNKTIVRFHSDSYTDITGFLAEYLSYDSSDPCPGQFTCRGCIKRLRCD 300
DB 413 RSQFVVTNSNKTIVRFHSDSYTDITGFLAEYLSYDSSDPCPGQFTCRGCIKRLRCD 472

QY 301 GMAUCTHSDDELNCSCDAGHGFCTCKNFKCKPLFWVCDSDVNDGDNDEGSCSPAQTFR 360
DB 473 GMAUCTHSDDELNCSCDAGHGFCTCKNFKCKPLFWVCDSDVNDGDNDEGSCSPAQTFR 532

QY 361 SNGKCLSKSQOQNGKDCGSDGDEASCPKXNVVVTCTKTYRCLNGLCLSKGNPECDKED 420
DB 533 SNGKCLSKSQOQNGKDCGSDGDEASCPKXNVVVTCTKTYRCLNGLCLSKGNPECDKED 592

QY 421 CSDDSDDEKDCCGRLRSFTROARVVGCTADDEGEMPMQVSLHALGQGHICGASLISPMVLV 480
DB 593 CSDDSDDEKDCCGRLRSFTROARVVGCTADDEGEMPMQVSLHALGQGHICGASLISPMVLV 652

QY 481 SAACHYIDDRGFRYSDFPTQMTAFGLHDQSORAPGVQERRLKRIISHPFNDFTFDYDI 540
DB 653 SAACHYIDDRGFRYSDFPTQMTAFGLHDQSORAPGVQERRLKRIISHPFNDFTFDYDI 712

QY 541 ALLELEKPAEYSSWVRPICLPDASHVFPAGKAIWVTGMGHTQYGGTGAALLQKGEIRVIN 600
DB 713 ALLELEKPAEYSSWVRPICLPDASHVFPAGKAIWVTGMGHTQYGGTGAALLQKGEIRVIN 772

QY 601 QTTCEMLLPQOITPRMNCVGLSGGVNSCGDSSGGLSVADGRIFPAGVAVSWGDCAQ 660

Db 773 QTTCEMLLPQOITPRMNCVGLSGVDSCGDSGGPLSSVEADGRIFQAGVSVWGDCQAQ 832
Qy 661 RNKPGVYTRLPFRDWIKENTGV 683
Db 833 RNKPGVYTRLPFRDWIKENTGV 855

RESULT 16

US-10-104-271-2
; Sequence 2, Application US/10104271
; Publication No. US20030181658A1
; GENERAL INFORMATION:
; APPLICANT: Madison, Edwin
; APPLICANT: Jiumn-Chern Yeh
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING SERINE PROTEASE CVP14, THE ENCOI
; TITLE OF INVENTION: POLYPEPTIDES AND METHODS BASED THEREON
; FILE REFERENCE: 24745-1614
; CURRENT APPLICATION NUMBER: US/10/104,271
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: 60/278,166
; PRIOR FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-104-271-2

Query Match 99.1%; Score 3754; DB 14; Length 855;
Best Local Similarity 99.3%; Pred. No. 6,4e-275;
Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 MAERVVMLEPPRARSLSKSVVTSVVAFPDTSKTVQRTQDSCSGFGLHARGVELMRFTTGG 60
Db 173 MAERVVMLEPPRARSLSKSVVTSVVAFPDTSKTVQRTQDSCSGFGLHARGVELMRFTTGG 232
Qy 61 FPDSPYPAHARCQWALRGDADSVSLTFPSFDLASCDERGSDDLVTYNTLSPMEPHALVQ 120
Db 233 FPDSPYPAHARCQWALRGDADSVSLTFPSFDLASCDERGSDDLVTYNTLSPMEPHALVQ 292
Qy 121 LCGTTPSYNLTFHSSQNVLLITLITNTERRHGFATFQOLPRMSSCGRLRKAQGTEN 180
Db 293 LCGTTPSYNLTFHSSQNVLLITLITNTERRHGFATFQOLPRMSSCGRLRKAQGTEN 352
Qy 181 SPYYPGHYPENIDCTNMIIEVPNNQHVVRPKFFYLLLEPRRACGTCRKYVEINGEKXCGE 240
Db 353 SPYYPGHYPENIDCTNMIIEVPNNQHVVRPKFFYLLLEPRRACGTCRKYVEINGEKXCGE 412
Qy 241 RSQFVVTSSNKKITVRFHSDQSYTDGTGFLAEVLSYDSDPCPGQFCTGRGCIKRELKCD 300
Db 413 RSQFVVTSSNKKITVRFHSDQSYTDGTGFLAEVLSYDSDPCPGQFCTGRGCIKRELKCD 472
Qy 301 GMADCTDHSDELNCSCDAGHQFTCKNFKCKPLFWVCDSVNDGDNDEGCGSCPAQTFRC 360
Db 473 GMADCTDHSDELNCSCDAGHQFTCKNFKCKPLFWVCDSVNDGDNDEGCGSCPAQTFRC 532
Qy 361 SNGKCLSKSQOQCKGKDDCGSDSDEASCPKXNVVTCYKHTYRCINGLCLSKGNPECDGKED 420
Db 533 SNGKCLSKSQOQCKGKDDCGSDSDEASCPKXNVVTCYKHTYRCINGLCLSKGNPECDGKED 592
Qy 421 CSDSDEKDCDGLRSFTROARVVGATDADGEGWPMQVSLHALGQGHICGASLISPMWLV 480
Db 593 CSDSDEKDCDGLRSFTROARVVGATDADGEGWPMQVSLHALGQGHICGASLISPMWLV 652
Qy 481 SAHACYIDDRGFYSPTQWTAFLGLHDQSORSAPEGVQERRLKRIISHPFNDFTFDYDI 540
Db 653 SAHACYIDDRGFYSPTQWTAFLGLHDQSORSAPEGVQERRLKRIISHPFNDFTFDYDI 712
Qy 541 ALLELEKPAEYSSWVRICLPDASHVPAGKAIWWTGHTQYGGTALILQKGEIRVIN 600
Db 713 ALLELEKPAEYSSWVRICLPDASHVPAGKAIWWTGHTQYGGTALILQKGEIRVIN 772

Qy 601 QTTCEMLLPQOITPRMNCVGLSGVDSCGDSGGPLSSVEADGRIFQAGVSVWGDCQAQ 660
Db 773 QTTCEMLLPQOITPRMNCVGLSGVDSCGDSGGPLSSVEADGRIFQAGVSVWGDCQAQ 832
Qy 661 RNKPGVYTRLPFRDWIKENTGV 683
Db 833 RNKPGVYTRLPFRDWIKENTGV 855

RESULT 17

US-10-147-211A-2
; Sequence 2, Application US/10147211A
; Publication No. US20030235900A1
; GENERAL INFORMATION:
; APPLICANT: Madison, Edwin
; APPLICANT: Yeh, Jiumn-Chern
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING A TRANSMEMBRANE SERINE PROTEASE 1
; TITLE OF INVENTION: ENCODED POLYPEPTIDES AND METHODS BASED THEREON
; FILE REFERENCE: 24745-1616
; CURRENT APPLICATION NUMBER: US/10/147,211A
; CURRENT FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/291,001
; PRIOR FILING DATE: 2001-05-14
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-147-211A-2

Query Match 99.1%; Score 3754; DB 14; Length 855;
Best Local Similarity 99.3%; Pred. No. 6,4e-275;
Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 MAERVVMLEPPRARSLSKSVVTSVVAFPDTSKTVQRTQDSCSGFGLHARGVELMRFTTGG 60
Db 173 MAERVVMLEPPRARSLSKSVVTSVVAFPDTSKTVQRTQDSCSGFGLHARGVELMRFTTGG 232
Qy 61 FPDSPYPAHARCQWALRGDADSVSLTFPSFDLASCDERGSDDLVTYNTLSPMEPHALVQ 120
Db 233 FPDSPYPAHARCQWALRGDADSVSLTFPSFDLASCDERGSDDLVTYNTLSPMEPHALVQ 292
Qy 121 LCGTTPSYNLTFHSSQNVLLITLITNTERRHGFATFQOLPRMSSCGRLRKAQGTEN 180
Db 293 LCGTTPSYNLTFHSSQNVLLITLITNTERRHGFATFQOLPRMSSCGRLRKAQGTEN 352
Qy 181 SPYYPGHYPENIDCTNMIIEVPNNQHVVRPKFFYLLLEPRRACGTCRKYVEINGEKXCGE 240
Db 353 SPYYPGHYPENIDCTNMIIEVPNNQHVVRPKFFYLLLEPRRACGTCRKYVEINGEKXCGE 412
Qy 241 RSQFVVTSSNKKITVRFHSDQSYTDGTGFLAEVLSYDSDPCPGQFCTGRGCIKRELKCD 300
Db 413 RSQFVVTSSNKKITVRFHSDQSYTDGTGFLAEVLSYDSDPCPGQFCTGRGCIKRELKCD 472
Qy 301 GMADCTDHSDELNCSCDAGHQFTCKNFKCKPLFWVCDSVNDGDNDEGCGSCPAQTFRC 360
Db 473 GMADCTDHSDELNCSCDAGHQFTCKNFKCKPLFWVCDSVNDGDNDEGCGSCPAQTFRC 532
Qy 361 SNGKCLSKSQOQCKGKDDCGSDSDEASCPKXNVVTCYKHTYRCINGLCLSKGNPECDGKED 420
Db 533 SNGKCLSKSQOQCKGKDDCGSDSDEASCPKXNVVTCYKHTYRCINGLCLSKGNPECDGKED 592
Qy 421 CSDSDEKDCDGLRSFTROARVVGATDADGEGWPMQVSLHALGQGHICGASLISPMWLV 480
Db 593 CSDSDEKDCDGLRSFTROARVVGATDADGEGWPMQVSLHALGQGHICGASLISPMWLV 652
Qy 481 SAHACYIDDRGFYSPTQWTAFLGLHDQSORSAPEGVQERRLKRIISHPFNDFTFDYDI 540
Db 653 SAHACYIDDRGFYSPTQWTAFLGLHDQSORSAPEGVQERRLKRIISHPFNDFTFDYDI 712
Qy 541 ALLELEKPAEYSSWVRICLPDASHVPAGKAIWWTGHTQYGGTALILQKGEIRVIN 600

Db 713 ALLELEKPAEYSSWVRPCLPDAASHVFPAGKAIWVTGMGHTQYGTGALLQKEIRVIN 772
Qy 601 OTTCENTLLPOQITPRMNCVGFSLSGVDSCCGDSGGPLSSVEADGRIFQAGVWSMGDCQAQ 660
Db 773 OTTCENTLLPOQITPRMNCVGFSLSGVDSCCGDSGGPLSSVEADGRIFQAGVWSMGDCQAQ 832
Qy 661 RNKGVYTRLPFRDWIKENTGV 683
Db 833 RNKGVYTRLPFRDWIKENTGV 855

RESULT 18

US-10-156-214A-2
; Sequence 2, Application US/10156214A
; Publication No. US20040001801A1
; GENERAL INFORMATION:
; APPLICANT: Edwin L. Madison
; APPLICANT: Joseph Edward Sempke
; APPLICANT: George P. Vlausk
; APPLICANT: Scott Jeffrey Kemp
; APPLICANT: Mallareddy Komandla
; APPLICANT: Daniel Vanna Slev
; TITLE OF INVENTION: Conjugates Activated By Cell Surface Proteases and Therapeutic Us
; FILE REFERENCE: 24745-1611
; CURRENT APPLICATION NUMBER: US/10/156,214A
; CURRENT FILING DATE: 2002-05-23
; NUMBER OF SEQ ID NOS: 611
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-156-214A-2

Query Match 99.1%; Score 3754; DB 15; Length 855;
Best Local Similarity 99.3%; Pred. No. 6.4e-275;
Matches 676; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 MAERVVMLPRPARSLKSFVVTSVVAFPTDSKTVOQTODNSCFGLHARGVELMRFTTPG 60
Db 173 MAERVVMLPRPARSLKSFVVTSVVAFPTDSKTVOQTODNSCFGLHARGVELMRFTTPG 232
Qy 61 PPDSPPYAHARCCQALGADADSVSLTFRSGFDLASCBERGSDLVYNTTSLSPMEHALVQ 120
Db 233 PPDSPPYAHARCCQALGADADSVSLTFRSGFDLASCBERGSDLVYNTTSLSPMEHALVQ 292
Qy 121 LCGTYPPSYNLTFFHSSQNVLLITLITNTERHHPGFATFPQLPRMSSCGGLRKAQGTFN 180
Db 293 LCGTYPPSYNLTFFHSSQNVLLITLITNTERHHPGFATFPQLPRMSSCGGLRKAQGTFN 352
Qy 181 SPYYPGHYPPNIDCTWNIIEVNNQHVYRPFKFFLLBPRACGTCPODYVEINKEKTCGE 240
Db 353 SPYYPGHYPPNIDCTWNIIEVNNQHVYRPFKFFLLBPRACGTCPODYVEINKEKTCGE 412
Qy 241 RSQPVVTSNSNKTIVRHSQSDSYTDGTGLAETLSYSSDPCPGQFTCRGTGCIKEIRCD 300
Db 413 RSQPVVTSNSNKTIVRHSQSDSYTDGTGLAETLSYSSDPCPGQFTCRGTGCIKEIRCD 472
Qy 301 GMADCTHSDBELNCSGAGHOFCTCKNFKCKPLFWVCSVNDGNSDDEGSCSPAQTFRC 360
Db 473 GMADCTHSDBELNCSGAGHOFCTCKNFKCKPLFWVCSVNDGNSDDEGSCSPAQTFRC 532
Qy 361 SNGKCLSKSQCCNGKDDCGSDDEASCPKVVVVTCTKHTYRCLNGLCLSKNPECDKED 420
Db 533 SNGKCLSKSQCCNGKDDCGSDDEASCPKVVVVTCTKHTYRCLNGLCLSKNPECDKED 592
Qy 421 CSDDSDKDDCDCCGRLSTTRQARVVGTDADGEMPMOYSLHALCOGHI CGASLISPMVLV 480
Db 593 CSDDSDKDDCDCCGRLSTTRQARVVGTDADGEMPMOYSLHALCOGHI CGASLISPMVLV 652
Qy 481 SAHHCYIDDRGFRSDPTQWTAFLGLHDGORSAPGVQERLRKRIISHPFNDPTFPYDI 540

Db 653 SAHHCYIDDRGFRSDPTQWTAFLGLHDGORSAPGVQERLRKRIISHPFNDPTFPYDI 712
Qy 541 ALLELEKPAEYSSWVRPCLPDAASHVFPAGKAIWVTGMGHTQYGTGALLQKEIRVIN 600
Db 713 ALLELEKPAEYSSWVRPCLPDAASHVFPAGKAIWVTGMGHTQYGTGALLQKEIRVIN 772
Qy 601 OTTCENTLLPOQITPRMNCVGFSLSGVDSCCGDSGGPLSSVEADGRIFQAGVWSMGDCQAQ 660
Db 773 OTTCENTLLPOQITPRMNCVGFSLSGVDSCCGDSGGPLSSVEADGRIFQAGVWSMGDCQAQ 832
Qy 661 RNKGVYTRLPFRDWIKENTGV 683
Db 833 RNKGVYTRLPFRDWIKENTGV 855

RESULT 19

US-10-072-012-352
; Sequence 352, Application US/10072012
; Publication No. US2004003493A1
; GENERAL INFORMATION:
; APPLICANT: Tchernev, Velizar
; APPLICANT: Spytek, Kimberly
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Patuturajan, Meera
; APPLICANT: Shinkets, Richard
; APPLICANT: Li, Li
; APPLICANT: Gangoli, Bsha
; APPLICANT: Padigaru, Muraidhara
; APPLICANT: Anderson, David W.
; APPLICANT: Rastelli, Luca
; APPLICANT: Miller, Charles E.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Coleman, Steven D.
; APPLICANT: Wolenc, Adam R.
; APPLICANT: Pena, Carol E. A
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Grose, William M.
; APPLICANT: Alsobrook II, John P.
; APPLICANT: Lepley, Denise M.
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Burgess, Catherine E.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-258
; CURRENT APPLICATION NUMBER: US/10/072,012
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 60/265,102
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/265,514
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,412
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,395
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/266,406
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/266,767
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/267,057
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/266,975
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/267,459
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 191
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 352
; LENGTH: 855

```

; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-072-012-352

Query Match      99.1%; Score 3754; DB 15; Length 855;
Best Local Similarity 99.3%; Pred. No. 6,4e-275;
Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAERVVMLPPRARSLSKSPVTSVVAFPPTDSKTQRTQDONSCEGFLHARGVELMRFTTPG 60
DB 173 MAERVVMLPPRARSLSKSPVTSVVAFPPTDSKTQRTQDONSCEGFLHARGVELMRFTTPG 232
QY 61 PPDSPYPAHARCQWALRGDADSVLSLTFRSFDLASCDERGSDLVYVNTLSPMEPHALVQ 120
DB 233 PPDSPYPAHARCQWALRGDADSVLSLTFRSFDLASCDERGSDLVYVNTLSPMEPHALVQ 292
QY 121 LCGTTPPSVNLTFHSSONVLLITLINTERRHHPGFATFPQLPRMSSCGRLRKAQGTN 180
DB 293 LCGTTPPSVNLTFHSSONVLLITLINTERRHHPGFATFPQLPRMSSCGRLRKAQGTN 352
QY 181 SPYYPGHYPNIDCTWNIIEVPNNQHVAFKFFYLLPBRACGTCPPDYVEINGEKYCGE 240
DB 353 SPYYPGHYPNIDCTWNIIEVPNNQHVAFKFFYLLPBRACGTCPPDYVEINGEKYCGE 412
QY 241 RSQFVVTSSNKTIVRFHSDQSYTDGTGFLAEVLSYSSDPCPGQFCRTGRICRKLRC 300
DB 413 RSQFVVTSSNKTIVRFHSDQSYTDGTGFLAEVLSYSSDPCPGQFCRTGRICRKLRC 472
QY 301 GMDCTHSDBELNCSDAHGQFTCKNFKCKPLFWVCDSVNDCGNSDBOGSCCPAQTFRC 360
DB 473 GMDCTHSDBELNCSDAHGQFTCKNFKCKPLFWVCDSVNDCGNSDBOGSCCPAQTFRC 532
QY 361 SNGKLSKSQOQCNKDCGSDGSDASCPKVVVVTCTGHTYRCNLGLCLSKGNPECDKED 420
DB 533 SNGKLSKSQOQCNKDCGSDGSDASCPKVVVVTCTGHTYRCNLGLCLSKGNPECDKED 592
QY 421 CSDGSDKDCDCGIRSTRQARVVGGTDADGEMPMQVSLHALGQGHICGASLISPMVLV 480
DB 593 CSDGSDKDCDCGIRSTRQARVVGGTDADGEMPMQVSLHALGQGHICGASLISPMVLV 652
QY 481 SAHACVYDDGFRSDPTQMTAFGLHDOSRSPAGVQERLKRITISHPPNDTFFYDI 540
DB 653 SAHACVYDDGFRSDPTQMTAFGLHDOSRSPAGVQERLKRITISHPPNDTFFYDI 712
QY 541 ALLELEKPAEYSSWVRICLPDASHVFPAGKAIWVTGHTQYGGTALLQKEIRVIN 600
DB 713 ALLELEKPAEYSSWVRICLPDASHVFPAGKAIWVTGHTQYGGTALLQKEIRVIN 772
QY 601 QTTCENILLPOQITPRMNCVGFSLSGVDSCQDGSGLPSSVADGRIFQAGVSWGDCAQ 660
DB 773 QTTCENILLPOQITPRMNCVGFSLSGVDSCQDGSGLPSSVADGRIFQAGVSWGDCAQ 832
QY 661 RNKRGVYTRLPLPDMWIKENTGV 683
DB 833 RNKRGVYTRLPLPDMWIKENTGV 855

RESULT 20
US-10-072-012-411
; Sequence 411, Application US/10072012
; Publication No. US2004003493A1
; GENERAL INFORMATION:
; APPLICANT: Tchernev, Velizar
; APPLICANT: Spytek, Kimberly
; APPLICANT: Zorhusen, Bryan
; APPLICANT: Patuturajan, Meera
; APPLICANT: Shmukets, Richard
; APPLICANT: Li, Li
; APPLICANT: Gangoli, Esha
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Anderson, David W.
; APPLICANT: Rastelli, Luca
; APPLICANT: Miller, Charles E.
```

```

; APPLICANT: Gerlach, Valerie
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Colman, Steven D.
; APPLICANT: Wolenc, Adam R.
; APPLICANT: Pena, Carol E. A
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Grosse, William M.
; APPLICANT: Alsobrook II, John P.
; APPLICANT: Lepley, Denise M.
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Burgess, Catherine E.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-258
; CURRENT APPLICATION NUMBER: US/10/072,012
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 60/265,102
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/265,514
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,412
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,395
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/266,406
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/266,767
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/267,057
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/266,975
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/267,459
; PRIOR FILING DATE: 2001-02-08
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1391
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 411
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-072-012-411

Query Match      99.1%; Score 3754; DB 15; Length 855;
Best Local Similarity 99.3%; Pred. No. 6,4e-275;
Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAERVVMLPPRARSLSKSPVTSVVAFPPTDSKTQRTQDONSCEGFLHARGVELMRFTTPG 60
DB 173 MAERVVMLPPRARSLSKSPVTSVVAFPPTDSKTQRTQDONSCEGFLHARGVELMRFTTPG 232
QY 61 PPDSPYPAHARCQWALRGDADSVLSLTFRSFDLASCDERGSDLVYVNTLSPMEPHALVQ 120
DB 233 PPDSPYPAHARCQWALRGDADSVLSLTFRSFDLASCDERGSDLVYVNTLSPMEPHALVQ 292
QY 121 LCGTTPPSVNLTFHSSONVLLITLINTERRHHPGFATFPQLPRMSSCGRLRKAQGTN 180
DB 293 LCGTTPPSVNLTFHSSONVLLITLINTERRHHPGFATFPQLPRMSSCGRLRKAQGTN 352
QY 181 SPYYPGHYPNIDCTWNIIEVPNNQHVAFKFFYLLPBRACGTCPPDYVEINGEKYCGE 240
DB 353 SPYYPGHYPNIDCTWNIIEVPNNQHVAFKFFYLLPBRACGTCPPDYVEINGEKYCGE 412
QY 241 RSQFVVTSSNKTIVRFHSDQSYTDGTGFLAEVLSYSSDPCPGQFCRTGRICRKLRC 300
DB 413 RSQFVVTSSNKTIVRFHSDQSYTDGTGFLAEVLSYSSDPCPGQFCRTGRICRKLRC 472
QY 301 GMDCTHSDBELNCSDAHGQFTCKNFKCKPLFWVCDSVNDCGNSDBOGSCCPAQTFRC 360
DB 473 GMDCTHSDBELNCSDAHGQFTCKNFKCKPLFWVCDSVNDCGNSDBOGSCCPAQTFRC 532
```

```
OY 361 SNGKCLSKSQOQNGKDGSDGSDASCPKXNVVTCYKHYRCLNGLCLSKGNPECDKED 420
|||
DB 533 SNGKCLSKSQOQNGKDGSDGSDASCPKXNVVTCYKHYRCLNGLCLSKGNPECDKED 592
OY 421 CSDGSDKDCDCGRLSFTROARVVGTDADGEMPMQVSLHALGQGHICGASLISPMNLV 480
CSDGSDKDCDCGRLSFTROARVVGTDADGEMPMQVSLHALGQGHICGASLISPMNLV 652
DB 593 CSDGSDKDCDCGRLSFTROARVVGTDADGEMPMQVSLHALGQGHICGASLISPMNLV 652
OY 481 SAACHYIDDRGFRSDPTQMTAFGLHDQSORAPGVQERLKRILISHPFNDPTFYDI 540
543 SAACHYIDDRGFRSDPTQMTAFGLHDQSORAPGVQERLKRILISHPFNDPTFYDI 712
OY 541 ALLELEKPAEYSSWVRPCLPDASHVPAGKAIWVTGMGHTQYGGTGALILQKEIRVIN 600
|||
DB 713 ALLELEKPAEYSSWVRPCLPDASHVPAGKAIWVTGMGHTQYGGTGALILQKEIRVIN 772
OY 601 QTTCENTLLPQOITPRMNCVGLSGGVDSGQDSSGGLSVADGRIFQAGVSWGDCQAQ 660
773 QTTCENTLLPQOITPRMNCVGLSGGVDSGQDSSGGLSVADGRIFQAGVSWGDCQAQ 832
OY 661 RNKRGVYTRLPFRDWMKENTGV 683
DB 833 RNKRGVYTRLPFRDWMKENTGV 855
```

```
RESULT 21
US-10-072-012-418
; Sequence 418, Application US/10072012
; Publication No. US2004003493A1
; GENERAL INFORMATION:
; APPLICANT: Tchernev, Velizar
; APPLICANT: Zetnusen, Bryan
; APPLICANT: Paturajan, Meera
; APPLICANT: Shimkets, Richard
; APPLICANT: Li, Li
; APPLICANT: Gangolli, Esna
; APPLICANT: Padigar, Muralidhara
; APPLICANT: Anderson, David W.
; APPLICANT: Rastelli, Luca
; APPLICANT: Miller, Charles E.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Gusev, Vladimr Y.
; APPLICANT: Coleman, Steven D.
; APPLICANT: Wolenc, Adam R.
; APPLICANT: Pena, Carol E. A
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Grose, William M.
; APPLICANT: Alsbrook II, John P.
; APPLICANT: Lepley, Denise M.
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Burgess, Catherine E.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-258
; CURRENT APPLICATION NUMBER: US/10/072.012
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 60/265,102
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/265,514
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,412
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,395
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/266,406
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/266,767
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/267,057
; PRIOR FILING DATE: 2001-02-07
```

```
; PRIOR APPLICATION NUMBER: 60/266,975
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/267,459
; PRIOR FILING DATE: 2001-02-08
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1391
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO: 418
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-072-012-418
```

```
Query Match 99.1%; Score 3754; DB 15; Length 855;
Best Local Similarity 99.3%; Pred. No. 6.4e-215;
Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
```

```
OY 1 MAERVYMLPPRARSLKSFVYTSVAPFPDSTKVQRTQDNSCSFGHLARGVELMRFTTPG 60
DB 173 MAERVYMLPPRARSLKSFVYTSVAPFPDSTKVQRTQDNSCSFGHLARGVELMRFTTPG 232
OY 61 FPDSPYAHARQWALRGDADSVLSLTFRSFDLASCDERGSIDLVTYNTLSPEPHALVQ 120
DB 233 FPDSPYAHARQWALRGDADSVLSLTFRSFDLASCDERGSIDLVTYNTLSPEPHALVQ 292
OY 121 LCGTTPSYNLTFHSSQNVLLITLITTERRHGFETFPOLBMSCGRLRKAQCTFN 180
DB 293 LCGTTPSYNLTFHSSQNVLLITLITTERRHGFETFPOLBMSCGRLRKAQCTFN 352
OY 181 SPYYPGHPNIDCTWNIIEVNNQHVVRPKFYLLEPRACGTCPCDYVEINKEKTCGE 240
DB 353 SPYYPGHPNIDCTWNIIEVNNQHVVRPKFYLLEPRACGTCPCDYVEINKEKTCGE 412
OY 241 RSQFVYTSNSNKTITRPHSDQSYTDGFLAEYLSYDSSDPCQOFTCTGRCTIRKELRCD 300
DB 413 RSQFVYTSNSNKTITRPHSDQSYTDGFLAEYLSYDSSDPCQOFTCTGRCTIRKELRCD 472
OY 301 GMADCTHSDLELNCSCDAGHOFCKKFKCPFLVNCVNDCCDNDSDGSCCPAOTFRC 360
DB 473 GMADCTHSDLELNCSCDAGHOFCKKFKCPFLVNCVNDCCDNDSDGSCCPAOTFRC 532
OY 361 SNGKCLSKSQOQNGKDGSDGSDASCPKXNVVTCYKHYRCLNGLCLSKGNPECDKED 420
DB 533 SNGKCLSKSQOQNGKDGSDGSDASCPKXNVVTCYKHYRCLNGLCLSKGNPECDKED 592
OY 421 CSDGSDKDCDCGRLSFTROARVVGTDADGEMPMQVSLHALGQGHICGASLISPMNLV 480
DB 593 CSDGSDKDCDCGRLSFTROARVVGTDADGEMPMQVSLHALGQGHICGASLISPMNLV 652
OY 481 SAACHYIDDRGFRSDPTQMTAFGLHDQSORAPGVQERLKRILISHPFNDPTFYDI 540
DB 543 SAACHYIDDRGFRSDPTQMTAFGLHDQSORAPGVQERLKRILISHPFNDPTFYDI 712
OY 541 ALLELEKPAEYSSWVRPCLPDASHVPAGKAIWVTGMGHTQYGGTGALILQKEIRVIN 600
DB 713 ALLELEKPAEYSSWVRPCLPDASHVPAGKAIWVTGMGHTQYGGTGALILQKEIRVIN 772
OY 601 QTTCENTLLPQOITPRMNCVGLSGGVDSGQDSSGGLSVADGRIFQAGVSWGDCQAQ 660
DB 773 QTTCENTLLPQOITPRMNCVGLSGGVDSGQDSSGGLSVADGRIFQAGVSWGDCQAQ 832
OY 661 RNKRGVYTRLPFRDWMKENTGV 683
DB 833 RNKRGVYTRLPFRDWMKENTGV 855
```

```
RESULT 22
US-10-600-187-2
; Sequence 2, Application US/10600187
; Publication No. US20040086910A1
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Tanimoto, Hirotooshi
```

```
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease
; FILE REFERENCE: D6064CIP/D
; CURRENT APPLICATION NUMBER: US/10/600,187
; CURRENT FILING DATE: 2003-06-20
; PRIOR APPLICATION NUMBER: US/09/654,600A
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 09/421,213
; 09/027,337
; PRIOR FILING DATE: 1999-10-20
; 1998-02-20
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: TADG-15
; US-10-600-187-2
```

```
Query Match          99.1%; Score 3754; DB 15; Length 855;
Best Local Similarity 99.3%; Pred. No. 6.4e-275;
Matches 676; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
```

```
QY 1 MAERVVMLPPRARSLKSPVVTSVVAFPTDSKTIVQRTQDNSCSFGIARGVLMRFTTPG 60
DB 173 MAERVVMLPPRARSLKSPVVTSVVAFPTDSKTIVQRTQDNSCSFGIARGVLMRFTTPG 232
QY 61 FPDSPYPAHARCOMALRGDADSVLSLTFRSFDLASCDERGSIDLVTYNTLSPEMHALVQ 120
DB 233 FPDSPYPAHARCOMALRGDADSVLSLTFRSFDLASCDERGSIDLVTYNTLSPEMHALVQ 292
QY 121 LCGTTPPSYNTLTFHSSQNVLLITLINTERRHGFATFQQLPRMSSCGRLRKAQGTN 180
DB 293 LCGTTPPSYNTLTFHSSQNVLLITLINTERRHGFATFQQLPRMSSCGRLRKAQGTN 352
QY 181 SPYYPGHYPNIDCTNMIIEVNNQHVKRFKFFYLLEPPRACGTCPKDYVEINKEKTCGE 240
DB 353 SPYYPGHYPNIDCTNMIIEVNNQHVKRFKFFYLLEPPRACGTCPKDYVEINKEKTCGE 412
QY 241 RSQGVVTSNSNKTIVTFHSDQSYTDGFLAEYLSYDSDPCPGQFTCRGRCIRKELRCD 300
DB 413 RSQGVVTSNSNKTIVTFHSDQSYTDGFLAEYLSYDSDPCPGQFTCRGRCIRKELRCD 472
QY 301 GMAICTDHSDELNCSCDAGHQFTCKNFKCKPLFWVCDVNDGDNDEGSCSPAQTFRC 360
DB 473 GMAICTDHSDELNCSCDAGHQFTCKNFKCKPLFWVCDVNDGDNDEGSCSPAQTFRC 532
QY 361 SNGKCLSKSQOQCKGKDDCGSDGDEASCPRKVVVTCYKHYTRCLNGCLSKGNPECDGKED 420
DB 533 SNGKCLSKSQOQCKGKDDCGSDGDEASCPRKVVVTCYKHYTRCLNGCLSKGNPECDGKED 592
QY 421 CSDGSDKDDCCGLRSTFRQARVVGCTDADEGEMQVSLHALGQGHICGASLISPMWLV 480
DB 593 CSDGSDKDDCCGLRSTFRQARVVGCTDADEGEMQVSLHALGQGHICGASLISPMWLV 652
QY 481 SAHACYIDRGFRYSPTQMTAFGLHDOSQSAAPGVQERRLRKRIISHPEFNDFTDYDI 540
DB 653 SAHACYIDRGFRYSPTQMTAFGLHDOSQSAAPGVQERRLRKRIISHPEFNDFTDYDI 712
QY 541 ALLELEKPAEYSSMWVRPCLPDASHVFPAGKAIWVTGMGHTQYGGTALILQKEIRVIN 600
DB 713 ALLELEKPAEYSSMWVRPCLPDASHVFPAGKAIWVTGMGHTQYGGTALILQKEIRVIN 772
QY 601 QTTCEMLPQOITPRMNCVGFISGVVSCQDSCGRLSVEADRIIFQAGVSWGDCAQ 660
DB 773 QTTCEMLPQOITPRMNCVGFISGVVSCQDSCGRLSVEADRIIFQAGVSWGDCAQ 832
QY 661 RNKRGVYTRLPLFRDWIKENTGV 683
DB 833 RNKRGVYTRLPLFRDWIKENTGV 855
```

```
RESULT 23
US-10-276-774-1798
; Sequence 1798, Application US/10276774
; Publication No. US20040053245A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; APPLICANT: Tang, Y, Tom et al
; TITLE OF INVENTION: No. US20040053245A1e1 Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-030
; CURRENT APPLICATION NUMBER: US/10/276,774
; CURRENT FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 2700
; SOFTWARE: Custom
; SEQ ID NO 1798
; LENGTH: 851
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-276-774-1798
```

```
Query Match          98.8%; Score 3745; DB 15; Length 851;
Best Local Similarity 99.0%; Pred. No. 3e-274;
Matches 676; Conservative 1; Mismatches 6; Indels 0; Gaps 0;
```

```
QY 1 MAERVVMLPPRARSLKSPVVTSVVAFPTDSKTIVQRTQDNSCSFGIARGVLMRFTTPG 60
DB 169 MAERVVMLPPRARSLKSPVVTSVVAFPTDSKTIVQRTQDNSCSFGIARGVLMRFTTPG 228
QY 61 FPDSPYPAHARCOMALRGDADSVLSLTFRSFDLASCDERGSIDLVTYNTLSPEMHALVQ 120
DB 229 FPDSPYPAHARCOMALRGDADSVLSLTFRSFDLASCDERGHILVTYNTLSPEMHALVQ 288
QY 121 LCGTTPPSYNTLTFHSSQNVLLITLINTERRHGFATFQQLPRMSSCGRLRKAQGTN 180
DB 289 LCGTTPPSYNTLTFHSSQNVLLITLINTERRHGFATFQQLPRMSSCGRLRKAQGTN 348
QY 181 SPYYPGHYPNIDCTNMIIEVNNQHVKRFKFFYLLEPPRACGTCPKDYVEINKEKTCGE 240
DB 349 SPYYPGHYPNIDCTNMIIEVNNQHVKRFKFFYLLEPPRACGTCPKDYVEINKEKTCGE 408
QY 241 RSQGVVTSNSNKTIVTFHSDQSYTDGFLAEYLSYDSDPCPGQFTCRGRCIRKELRCD 300
DB 409 RSQGVVTSNSNKTIVTFHSDQSYTDGFLAEYLSYDSDPCPGQFTCRGRCIRKELRCD 468
QY 301 GMAICTDHSDELNCSCDAGHQFTCKNFKCKPLFWVCDVNDGDNDEGSCSPAQTFRC 360
DB 469 GMAICTDHSDELNCSCDAGHQFTCKNFKCKPLFWVCDVNDGDNDEGSCSPAQTFRC 528
QY 361 SNGKCLSKSQOQCKGKDDCGSDGDEASCPRKVVVTCYKHYTRCLNGCLSKGNPECDGKED 420
DB 529 SNGKCLSKSQOQCKGKDDCGSDGDEASCPRKVVVTCYKHYTRCLNGCLSKGNPECDGKED 588
QY 421 CSDGSDKDDCCGLRSTFRQARVVGCTDADEGEMQVSLHALGQGHICGASLISPMWLV 480
DB 589 CSDGSDKDDCCGLRSTFRQARVVGCTDADEGEMQVSLHALGQGHICGASLISPMWLV 648
QY 481 SAHACYIDRGFRYSPTQMTAFGLHDOSQSAAPGVQERRLRKRIISHPEFNDFTDYDI 540
DB 649 SAHACYIDRGFRYSPTQMTAFGLHDOSQSAAPGVQERRLRKRIISHPEFNDFTDYDI 708
QY 541 ALLELEKPAEYSSMWVRPCLPDASHVFPAGKAIWVTGMGHTQYGGTALILQKEIRVIN 600
DB 709 ALLELEKPAEYSSMWVRPCLPDASHVFPAGKAIWVTGMGHTQYGGTALILQKEIRVIN 768
QY 601 QTTCEMLPQOITPRMNCVGFISGVVSCQDSCGRLSVEADRIIFQAGVSWGDCAQ 660
DB 769 QTTCEMLPQOITPRMNCVGFISGVVSCQDSCGRLSVEADRIIFQAGVSWGDCAQ 828
QY 661 RNKRGVYTRLPLFRDWIKENTGV 683
```

Db 829 RNKPGVYTRLPFRDWIKENTGV 851

```
RESULT 24
US-10-296-115-1143
; Sequence 1143, Application US/10296115
; Publication No. US20040053248A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq Inc
; TITLE OF INVENTION: No. US20040053248A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 784PCT
; CURRENT APPLICATION NUMBER: US/10/296.115
; CURRENT FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: US09/488,725
; PRIOR FILING DATE: 2000-01-22
; PRIOR APPLICATION NUMBER: US09/552,317
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 1478
; SEQ ID NO 1143
; LENGTH: 851
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-296-115-1143
```

Query Match 98.8%; Score 3745; DB 15; Length 851;
Best Local Similarity 99.0%; Pred. No. 3e-274;
Matches 676; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

```
Qy 1 MAERVVMLPPRARSLSFVYTSVVAFPDTSKYVQRTQDNCSEFGLHARGVELMRFTTPG 60
Db 169 MAERVVMLPPRARSLSFVYTSVVAFPDTSKYVQRTQDNCSEFGLHARGVELMRFTTPG 228
Qy 61 PPDSPPYAHARCOMALRGDADSVLSLFRSFDLASCDERSDLYTVYNTLSPMEPHALVQ 120
Db 229 PPDSPPYAHARCOMALRGDADSVLSLFRSFDLASCDERHMLTVYNTLSPMEPHALVQ 288
Qy 121 LCGTYPSPYNTLFRSSQNVLLITLITNTERRHGFEATFQOLPRMSSCGRLRKAQGTFN 180
Db 289 LCGTYPSPYNTLFRSSQNVLLITLITNTERRHGFEATFQOLPRMSSCGRLRKAQGTFN 348
Qy 181 SPYYPGHYPNIDCTWNIIEVNNQHVKKRFKFLLEPRACGTCPKDYVEINKEKCGE 240
Db 349 SPYYPGHYPNIDCTWNIIEVNNQHVKKRFKFLLEPRAGVAGTCPKDYVEINKEKCGE 408
Qy 241 RSQFVTVSNKTIIVRFRHSQSYTDGFLAFLYLSYSSDPCPGQFTCTGCRIRKELRCD 300
Db 409 RSQFVTVSNKTIIVRFRHSQSYTDGFLAFLYLSYSSDPCPGQFTCTGCRIRKELRCD 468
Qy 301 GMADCTHSDDELNCSGAGHOFCKNFKCPLFWVCDSVNDCCGNSDEGSCSPAOPTFC 360
Db 469 GMADCTHSDDELNCSGAGHOFCKNFKCPLFWVCDSVNDCCGNSDEGSCSPAOPTFC 528
Qy 361 SNGKCLSKSQCGNCKDCCGDSDEASCPKVVVVTCTKTYRCLNGLCLSKNPECDGKED 420
Db 529 SNGKCLSKSQCGNCKDCCGDSDEASCPKVVVVTCTKTYRCLNGLCLSKNPECDGKED 588
Qy 421 CSDGSEKDCDCGRLSFTROARVVGCTDADGEMPMOVSLLHALCGHICGASLISPMVLV 480
Db 589 CSDGSEKDCDCGRLSFTROARVVGCTDADGEMPMOVSLLHALCGHICGASLISPMVLV 648
Qy 481 SAAHCYIDDRGFRSDPTQWTAFLGIDHOSORSAFGVERLKRILISHPFNDTTFDDI 540
Db 649 SAAHCYIDDRGFRSDPTQWTAFLGIDHOSORSAFGVERLKRILISHPFNDTTFDDI 708
Qy 541 ALLELEKPAEYSNWRPILCPDASHVPPAGKAIWVTGHTQYGGTGAALLQKEIRVIN 600
Db 709 ALLELEKPAEYSNWRPILCPDASHVPPAGKAIWVTGHTQYGGTGAALLQKEIRVIN 768
Qy 601 QTTCEMLLPQOITPRMNCVGLSGVDSCGDSGGGLSSVEADGRIFQAGVYSGDGAQ 660
Db 769 QTTCEMLLPQOITPRMNCVGLSGVDSCGDSGGGLSSVEADGRIFQAGVYSGDGAQ 828
Qy 661 RNKPGVYTRLPFRDWIKENTGV 683
```

Db 829 RNKPGVYTRLPFRDWIKENTGV 851

```
RESULT 25
US-10-097-340-312
; Sequence 312, Application US/10097340
; Publication No. US20030087250A1
; GENERAL INFORMATION:
; APPLICANT: John MONAHAN
; APPLICANT: Manjula GANNANARAPU
; APPLICANT: Sebastian HOERSCH
; APPLICANT: Shubhangi KAMATKAR
; APPLICANT: Steve G. KOVATS
; APPLICANT: Rachel E. MEYERS
; APPLICANT: Michael MORRISSEY
; APPLICANT: Peter OLANDT
; APPLICANT: Ami SEN
; APPLICANT: Peter VEIBY
; APPLICANT: Gordon B. MILLS
; APPLICANT: Robert C. BAST, Jr.
; APPLICANT: Karen LU
; APPLICANT: Rosemarie SCHMANDT
; APPLICANT: Xumei ZHAO
; APPLICANT: Karen GLATT
; TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,
; TITLE OF INVENTION: Assessment, Prevention, and Therapy of Ovarian Cancer
; FILE REFERENCE: MRI-030
; CURRENT APPLICATION NUMBER: US/10/097,340
; CURRENT FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: 60/276,025
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/325,149
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/276,026
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/324,967
; PRIOR FILING DATE: 2001/09/26
; PRIOR APPLICATION NUMBER: 60/311,732
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/325,102
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/323,580
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 363
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 312
; LENGTH: 782
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-097-340-312
```

Query Match 86.0%; Score 3259.5; DB 14; Length 782;
Best Local Similarity 88.4%; Pred. No. 1.3e-237;
Matches 604; Conservative 1; Mismatches 5; Indels 73; Gaps 2;

```
Qy 1 MAERVVMLPPRARSLSFVYTSVVAFPDTSKYVQRTQDNCSEFGLHARGVELMRFTTPG 60
Db 173 MAERVVMLPPRARSLSFVYTSVVAFPDTSKYVQRTQDNCSEFGLHARGVELMRFTTPG 232
Qy 61 PPDSPPYAHARCOMALRGDADSVLSLFRSFDLASCDERSDLYTVYNTLSPMEPHALVQ 120
Db 233 PPDSPPYAHARCOMALRGDADSVLSLFRSFDLASCDERSDLYTVYNTLSPMEPHALVQ 292
Qy 121 LCGTYPSPYNTLFRSSQNVLLITLITNTERRHGFEATFQOLPRMSSCGRLRKAQGTFN 180
Db 293 LCGTYPSPYNTLFRSSQNVLLITLITNTERRHGFEATFQOLPRMSSCGRLRKAQGTFN 352
Qy 181 SPYYPGHYPNIDCTWNIIEVNNQHVKKRFKFLLEPRACGTCPKDYVEINKEKCGE 240
Db 353 SPYYPGHYPNIDCTWNIIEVNNQHVKKRFKFLLEPRAGVAGTCPKDYVEINKEKCGE 412
Qy 241 RSQFVTVSNKTIIVRFRHSQSYTDGFLAFLYLSYSSDPCPGQFTCTGCRIRKELRCD 300
```

```
Db 413 RSQFVVTNSNKTITVRHSDQSYDTGTGFLAEYLSYDSSD----- 451
Qy 301 GMAADCTDHSDELINSCSAGHOFCTCKNKECPLEFVWCDNVNDCGNSDEOGSCPAOTFR 360
Db 452 -----RCAGHOFCTCKNKECPLEFVWCDNVNDCGNSDEOGC----- 488
Qy 361 SNGKCLSKSQOQNGKXDCGSDGSDASCPKVVVVTCTKHTYRCINGLCLSKGNPECDKED 420
Db 489 -----MNVVVTCTGHTYRCINGLCLSKGNPECDKED 519
Qy 421 CSQGSDEKDCDCGLRSFTQARVVGCTDADBEWPMQVSLHALGQGHICGASLISPMWLV 480
Db 520 CSQGSDEKDCDCGLRSFTQARVVGCTDADBEWPMQVSLHALGQGHICGASLISPMWLV 579
Qy 481 SAACHYIDRGRFYSPTQMTAFGLHDOSRSPAGVQERRLKRIISHPFNDFTFDYDI 540
Db 580 SAACHYIDRGRFYSPTQMTAFGLHDOSRSPAGVQERRLKRIISHPFNDFTFDYDI 639
Qy 541 ALLELEKPAEYSSWVRPICI.PDASHVFPAGKAIWVTGMHTQYGGTALILQKEIRVIN 600
Db 640 ALLELEKPAEYSSWVRPICI.PDASHVFPAGKAIWVTGMHTQYGGTALILQKEIRVIN 699
Qy 601 QTTCEMLPQOITPRMNCVGLSGVSDSCQDSGGPLSYBADGRITQAGVSWGDCQAQ 660
Db 700 QTTCEMLPQOITPRMNCVGLSGVSDSCQDSGGPLSYBADGRITQAGVSWGDCQAQ 759
Qy 661 RNKGVYTRLP.LFRDWIKENTGV 683
Db 760 RNKGVYTRLP.LFRDWIKENTGV 782
```

```
RESULT 26
US-09-900-751-2
; Sequence 2, Application US/09900751
; Patent No. US2002026653A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Keith D
; APPLICANT: Leviten, Michael W.
; TITLE OF INVENTION: TRANSGENIC MICE CONTAINING SERINE
; TITLE OF INVENTION: PROTEASE GENE DISRUPTIONS
; FILE REFERENCE: R-386
; CURRENT APPLICATION NUMBER: US/09/900, 751
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US 60/217,449
; PRIOR FILING DATE: 2000-07-10
; PRIOR APPLICATION NUMBER: US 60/223,170
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: US 60/223,460
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-900-751-2
```

Query Match 84.7%; Score 3209; DB 9; Length 855;

Best Local Similarity 83.2%; Pred. No. 9,7e-234; Matches 568; Conservative 59; Mismatches 56; Indels 0; Gaps 0;

```
Qy 1 MAERVVMLEPPRARSLKSFVTVSVAPPTDSKTQVQRQDNSCSFGILHARGVELMRFPTPG 60
Db 173 MAERVVMLEPPRARSLKSFVTVSVAPPTDSKTQVQRQDNSCSFGILHARGVELMRFPTPG 232
Qy 61 PPSPPYAHARCOMALRGDADSVLSLTFRSFDLASCDERGSIDLVTYNTLSPEMHALVQ 120
Db 233 PPSPPYAHARCOMALRGDADSVLSLTFRSFDVASPCDEHGSIDLVTYNTLSPEMHALVYR 292
Qy 121 LCGTSPSYNLTTFHSSGNVLLTITNTTERRHRRGFELTFQOLPRMSSCGRLKKAQCTEN 180
Db 293 LCGTSPSYNLTTFHSSGNVLLTITNTTERRHRRGFELTFQOLPRMSSCGRLKKAQCTEN 352
```

```
Qy 181 SPYPCGHYPENIDCTNNIEVPNNQHVKVRFKFFYLEPRRACGTCRKYVEINGEKXCGE 240
Db 353 SPYPCGHYPENIDCTNNIEVPNNQHVKVRFKFFYLEPRRACGTCRKYVEINGEKXCGE 412
Qy 241 RSQFVVTNSNKTITVRHSDQSYDTGTGFLAEYLSYDSSDPCQOFTCRGCRCKELRCD 300
Db 413 RSQFVVTNSNKTITVRHSDQSYDTGTGFLAEYLSYDSSDPCQOFTCRGCRCKELRCD 472
Qy 301 GMAADCTDHSDELINSCSAGHOFCTCKNKECPLEFVWCDNVNDCGNSDEOGSCPAOTFR 360
Db 473 GMAADCTDHSDELINSCSAGHOFCTCKNKECPLEFVWCDNVNDCGNSDEOGSCPAOTFR 532
Qy 361 SNGKCLSKSQOQNGKXDCGSDGSDASCPKVVVVTCTKHTYRCINGLCLSKGNPECDKED 420
Db 533 SNGKCLPQSGKXGKXDCGSDGSDASCPKVVVVTCTKHTYRCINGLCLSKGNPECDKED 592
Qy 421 CSQGSDEKDCDCGLRSFTQARVVGCTDADBEWPMQVSLHALGQGHICGASLISPMWLV 480
Db 593 CSQGSDEKDCDCGLRSFTQARVVGCTDADBEWPMQVSLHALGQGHICGASLISPMWLV 652
Qy 481 SAACHYIDRGRFYSPTQMTAFGLHDOSRSPAGVQERRLKRIISHPFNDFTFDYDI 540
Db 653 SAACHYIDRGRFYSPTQMTAFGLHDOSRSPAGVQERRLKRIISHPFNDFTFDYDI 712
Qy 541 ALLELEKPAEYSSWVRPICI.PDASHVFPAGKAIWVTGMHTQYGGTALILQKEIRVIN 600
Db 713 ALLELEKPAEYSSWVRPICI.PDASHVFPAGKAIWVTGMHTQYGGTALILQKEIRVIN 772
Qy 601 QTTCEMLPQOITPRMNCVGLSGVSDSCQDSGGPLSYBADGRITQAGVSWGDCQAQ 660
Db 773 QTTCEMLPQOITPRMNCVGLSGVSDSCQDSGGPLSYBADGRITQAGVSWGDCQAQ 832
Qy 661 RNKGVYTRLP.LFRDWIKENTGV 683
Db 833 RNKGVYTRLP.LFRDWIKENTGV 855
```

```
RESULT 27
US-10-072-012-355
; Sequence 355, Application US/10072012
; Publication No. US2004003493A1
; GENERAL INFORMATION:
; APPLICANT: Tcherenev, Velizar
; APPLICANT: Spytek, Kimberly
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Patnursajan, Meera
; APPLICANT: Shimkete, Richard
; APPLICANT: Li, Li
; APPLICANT: Gangolli, Baha
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Anderson, David W.
; APPLICANT: Rastelli, Luca
; APPLICANT: Miller, Charles E.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Gusev, Vladimiro Y.
; APPLICANT: Colman, Steven D.
; APPLICANT: Wolenc, Adam R.
; APPLICANT: Pena, Carol E. A
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Grose, William M.
; APPLICANT: Alsobrook II, John P.
; APPLICANT: Lepley, Denise M.
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Burgess, Catherine E.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-258
; CURRENT APPLICATION NUMBER: US/10/072,012
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 60/265,102
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/265,514
```



```

; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,412
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,395
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/266,406
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/266,767
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/267,057
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/266,975
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/267,459
; PRIOR FILING DATE: 2001-02-08
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1391
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 355
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-072-012-355

Query Match      84.7%; Score 3209; DB 15; Length 855;
Best Local Similarity 83.2%; Pred. No. 9.7e-234;
Matches 568; Conservative 59; Mismatches 56; Indels 0; Gaps 0;

Qy 1 MAEERVMTLPPRARSLKSPVTVSVAPPTDSKTVOPTODNSCSFGLHARGVELMRFTTPG 60
Db 173 MAEERVMTLPPRARSLKSPVTVSVAPPTDSKTVOPTODNSCSFGLHARGVELMRFTTPG 232
Qy 61 FDPSPYAHARQWALRGDADSVLSTFRSFDLASCDESGSDLYTVYNTLSMPMHPALVQ 120
Db 233 FPNSPYAHARQWALRGDADSVLSTFRSFDVAPCEHSGSDLYTVYNTLSMPMHPALVVR 292
Qy 121 LCGTYPSPSYNLTFHSSQNVLLITLITNERRHPEPEATFQPLPMSSCGGRRLRAQGTEN 180
Db 293 LCGTFSPSYNLTFHSSQNVFLVTLITNDRHPGEATFQPLPMSSCGGFLSTPTQGTFS 352
Qy 181 SPYYPGHPYPRINDCTNMIENPNNGHVVYRPFKFLYLLBPRACGCPDYVEINEKXCGE 240
Db 353 SPYYPGHPYPRINDCTNMIENPNNGHVVYRPFKFLYLLBPRACGCPDYVEINEKXCGE 412
Qy 241 RSQFVTVSNENKITYRFRHSQSYDTGTFLAEYLSYDSSDPCPGQFTGRTGRCIRKELRCD 300
Db 413 RSQFVTVSNENKITYRFRHSQSYDTGTFLAEYLSYDSSDPCPGQFTGRTGRCIRKELRCD 472
Qy 301 GMADCTHSDLELNCSCDAHQFTCKNKFCKPLFVVCDSVNDGNSDBQGCCCPAQTRFC 360
Db 473 GMADCTHSDLELNCSCDAHQFTCKNKFCKPLFVVCDSVNDGNSDBQGCCCPAQTRFC 532
Qy 473 GMADCTHSDLELNCSCDAHQFTCKNKFCKPLFVVCDSVNDGNSDBQGCCCPAQTRFC 532
Db 473 GMADCTHSDLELNCSCDAHQFTCKNKFCKPLFVVCDSVNDGNSDBQGCCCPAQTRFC 532
Qy 361 SNGCCLSKSQOQCKNDKCGGSDASCPKXNVVYCTHTYRCLNGLCLSKNPECDKED 420
Db 533 SNGCCLSKSQOQCKNDKCGGSDASCPKXNVVYCTHTYRCLNGLCLSKNPECDKED 592
Qy 533 SNGCCLSKSQOQCKNDKCGGSDASCPKXNVVYCTHTYRCLNGLCLSKNPECDKED 592
Db 533 SNGCCLSKSQOQCKNDKCGGSDASCPKXNVVYCTHTYRCLNGLCLSKNPECDKED 592
Qy 421 CSDSDEKDCDCGARSFTQRARVVGCTDADGEMPMQVSLHALQGHICGASLSPMWLV 480
Db 593 CSDSDEKDCDCGARSFTQRARVVGCTDADGEMPMQVSLHALQGHICGASLSPMWLV 652
Qy 593 CSDSDEKDCDCGARSFTQRARVVGCTDADGEMPMQVSLHALQGHICGASLSPMWLV 652
Db 593 CSDSDEKDCDCGARSFTQRARVVGCTDADGEMPMQVSLHALQGHICGASLSPMWLV 652
Qy 481 SAHACVYDDRGFRYSPTQWTAFLGLHDQSORSAFGVQERRLKRIISHPFNDFTFYDI 540
Db 653 SAHACVYDDRGFRYSPTQWTAFLGLHDQSORSAFGVQERRLKRIISHPFNDFTFYDI 712
Qy 653 SAHACVYDDRGFRYSPTQWTAFLGLHDQSORSAFGVQERRLKRIISHPFNDFTFYDI 712
Db 653 SAHACVYDDRGFRYSPTQWTAFLGLHDQSORSAFGVQERRLKRIISHPFNDFTFYDI 712
Qy 541 ALLELEKPAEYSSNVVPTCLBDASHVPAGKAIWVTGNGHYOGTGLILQKGEIRVIN 600
Db 713 ALLELEKPAEYSSNVVPTCLBDASHVPAGKAIWVTGNGHYOGTGLILQKGEIRVIN 772
Qy 713 ALLELEKPAEYSSNVVPTCLBDASHVPAGKAIWVTGNGHYOGTGLILQKGEIRVIN 772
Db 713 ALLELEKPAEYSSNVVPTCLBDASHVPAGKAIWVTGNGHYOGTGLILQKGEIRVIN 772
Qy 601 QTTCEMLLPQOITPRMNCVGFSLSGVNSCQDSSGAPLSVBAQRIFQAGVNSWGDCAQ 660
Db 773 QTTCEMLLPQOITPRMNCVGFSLSGVNSCQDSSGAPLSVBAQRIFQAGVNSWGDCAQ 832
```

```

Qy 661 RNKGVYTRLPFRDMIXENTGV 683
Db 833 RNKGVYTRLPFRDMIXENTGV 855

RESULT 28
US-10-072-012-413
; Sequence 413, Application US/10072012
; Publication No. US2004003493A1
; GENERAL INFORMATION:
; APPLICANT: Tchernov, Velizar
; APPLICANT: Spytak, Kimberly
; APPLICANT: Zethusen, Bryan
; APPLICANT: Paturajan, Meera
; APPLICANT: Shukets, Richard
; APPLICANT: Li, Li
; APPLICANT: Gangoli, Baha
; APPLICANT: Padigaru, Muraidhara
; APPLICANT: Anderson, David W.
; APPLICANT: Rastelli, Luca
; APPLICANT: Miller, Charles E.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Coleman, Steven D.
; APPLICANT: Wolenc, Adam R.
; APPLICANT: Pena, Carol E. A
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Grose, William M.
; APPLICANT: Alsobrook II, John P.
; APPLICANT: Lepley, Denise M.
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Burgess, Catherine E.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-258
; CURRENT APPLICATION NUMBER: US/10/072,012
; PRIOR FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 60/265,102
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/265,514
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,412
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,395
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/266,406
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/266,767
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/267,057
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/266,975
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/267,459
; PRIOR FILING DATE: 2001-02-08
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1391
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 413
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-072-012-413

Query Match      84.7%; Score 3209; DB 15; Length 855;
Best Local Similarity 83.2%; Pred. No. 9.7e-234;
Matches 568; Conservative 59; Mismatches 56; Indels 0; Gaps 0;

Qy 1 MAEERVMTLPPRARSLKSPVTVSVAPPTDSKTVOPTODNSCSFGLHARGVELMRFTTPG 60
Db 173 MAEERVMTLPPRARSLKSPVTVSVAPPTDSKTVOPTODNSCSFGLHARGVELMRFTTPG 232
```

```
Db 173 MAVERVTLPPRARALKSFVLTSVVAFPIDPRMLQRTQDNSCSFALHAGAAVTRFTTPG 232
Qy 61 FDPSPYPAHARCQWALRGDADSVLSTFRSFDLASCDERGSIDLVTYNTLSPEMHALVQ 120
Db 233 FPNSPYPAHARCQWALRGDADSVLSTFRSFDVAPCDHEGSDLVYVYDLSLPEMHALVVR 292
Qy 121 LCGTFPSYVLTFRHSQNVLLITLITNTERRHGFEATFQLPMSGCGRLKKAQGTEN 180
Db 293 LCGTFPSYVLTFRHSQNVFLVTLITNTERRHGFEATFQLPMSGCGRLSPDTQGTFS 352
Qy 181 SPYPGHYPPNIDCTNNIEVPNNQHVVRFPFYLLEPRRACGTCPDYVEINEKXCGE 240
Db 353 SPYPGHYPPNIDCTNNIKVNNNANVVRFPFLFLVDPNVVPSCTQDYVEINEKXCGE 412
Qy 241 RSQFVVTNSNKTIVRFHSDQSYTDGTFLAEVLSYDSDPCPGQFTCRGTCIRKELRCD 300
Db 413 RSQFVVTNSNKTIVRHSHDSYTDGTFLAEVLSYDSDNDPCPGMFMCKTGRCIRKELRCD 472
Qy 301 GMADCTHSDDELNCSDAHQFTCKNRCFKPLFVWCDVNDCGNSDPQSCGCPAQTRFC 360
Db 473 GMADCPYSDERYCRCAHQFTCKNQFCCKPLFVWCDVNDCGDSDDEGSCSPAGSFKC 532
Qy 361 SNGKCLSKSQOQNGKDDCGSDSASCPKVVNVVTCGHTYRCUNGLCLSKGNPECCKED 420
Db 533 SNGKCLPQSQCKNGKDDCGSDSDASCDVNVVSCYKTYRCQNGCLSKGNPECCKED 592
Qy 421 CSDGSDKDDCGLSRFTQARVVGTDADEGEMPMQVSLHALGQGHICGASLISPMVLV 480
Db 593 CSDGSDKDDCGLSRFTKQARVVGTGTMADGEMPMQVSLHALGQGHICGASLISPMVLV 652
Qy 481 SAHCYTDGRGFRYSPTQMTAFGLHDOSORAPVQVERRLKRIISHPPNDFTFDYDI 540
Db 653 SAHCYTDGRGFRYSPTQMTAFGLHDOSORASGVQELKRIITHPSFNDTFDYDI 712
Qy 541 ALLELEKAPYSSMVRPCLPDASHFPAKAIWVTMGHTOVGSGTALILQKEIRVIN 600
Db 713 ALLELESVEYSIVVRICLPDATHVPAGALWVTMGHTKGSTGALILQKEIRVIN 772
Qy 601 QTTCEMLIPQOITPRMNCVGLSGVNSCGDSCGRLSYEADGRIFQAGVNSGDCQAQ 660
Db 773 QTTCEMLIPQOITPRMNCVGLSGVNSCGDSCGRLSYEADGRIFQAGVNSGDCQAQ 832
Qy 661 RNKRGVYTRLPFRDWIKENTGV 683
Db 833 RNKRGVYTRLPVVRDWIKENTGV 855

RESULT 29
US-10-072-012-356
; Sequence 356, Application US/10072012
; Publication No. US2004003493A1
; GENERAL INFORMATION:
; APPLICANT: Tchernov, Velizar
; APPLICANT: Szytek, Kimberly
; APPLICANT: Zetnusen, Bryan
; APPLICANT: Paturajan, Meera
; APPLICANT: Shukets, Richard
; APPLICANT: Li, Li
; APPLICANT: Gangoli, Beha
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Anderson, David W.
; APPLICANT: Rastelli, Luca
; APPLICANT: Miller, Charles E.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Tauplier Jr, Raymond J.
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Colman, Steven D.
; APPLICANT: Wolenc, Adam R.
; APPLICANT: Pena, Carol E. A
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Grose, William M.
; APPLICANT: Alsobrook II, John P.
; APPLICANT: Lepley, Denise M.
```

```
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Burgess, Catherine E.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-258
; CURRENT APPLICATION NUMBER: US/10/072, 012
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 60/265,102
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/265,514
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,412
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,395
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/266,406
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/266,767
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/267,057
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/266,975
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/267,459
; PRIOR FILING DATE: 2001-02-08
; Remaining Prior Application data removed - See file wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1391
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 356
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-10-072-012-356

Query Match 84.4%; Score 3197; DB 15; Length 855;
Best Local Similarity 83.0%; Pred. No.7.8e-233;
Matches 567; Conservative 59; Mismatches 57; Indels 0; Gaps 0;

Qy 1 MAERVYMLPPRARLSKSFVYTSVVAFPIDPRMLQRTQDNSCSFALHAGAAVTRFTTPG 60
Db 173 MAVERVTLPPRARALKSFVLTSVVAFPIDPRMLQRTQDNSCSFALHAGAAVTRFTTPG 232
Qy 61 FDPSPYPAHARCQWALRGDADSVLSTFRSFDLASCDERGSIDLVTYNTLSPEMHALVQ 120
Db 233 FPNSPYPAHARCQWALRGDADSVLSTFRSFDVAPCDHEGSDLVYVYDLSLPEMHALVVR 292
Qy 121 LCGTFPSYVLTFRHSQNVLLITLITNTERRHGFEATFQLPMSGCGRLKKAQGTEN 180
Db 293 LCGTFPSYVLTFRHSQNVFLVTLITNTERRHGFEATFQLPMSGCGRLSPDTQGTFS 352
Qy 181 SPYPGHYPPNIDCTNNIEVPNNQHVVRFPFYLLEPRRACGTCPDYVEINEKXCGE 240
Db 353 SPYPGHYPPNIDCTNNIKVNNNANVVRFPFLFLVDPNVVPSCTQDYVEINEKXCGE 412
Qy 241 RSQFVVTNSNKTIVRFHSDQSYTDGTFLAEVLSYDSDPCPGQFTCRGTCIRKELRCD 300
Db 413 RSQFVVTNSNKTIVRHSHDSYTDGTFLAEVLSYDSDNDPCPGMFMCKTGRCIRKELRCD 472
Qy 301 GMADCTHSDDELNCSDAHQFTCKNRCFKPLFVWCDVNDCGNSDPQSCGCPAQTRFC 360
Db 473 GMADCPYSDERYCRCAHQFTCKNQFCCKPLFVWCDVNDCGDSDDEGSCSPAGSFKC 532
Qy 361 SNGKCLSKSQOQNGKDDCGSDSASCPKVVNVVTCGHTYRCUNGLCLSKGNPECCKED 420
Db 533 SNGKCLPQSQCKNGKDDCGSDSDASCDVNVVSCYKTYRCQNGCLSKGNPECCKED 592
Qy 421 CSDGSDKDDCGLSRFTQARVVGTDADEGEMPMQVSLHALGQGHICGASLISPMVLV 480
Db 593 CSDGSDKDDCGLSRFTKQARVVGTGTMADGEMPMQVSLHALGQGHICGASLISPMVLV 652
Qy 481 SAHCYTDGRGFRYSPTQMTAFGLHDOSORAPVQVERRLKRIISHPPNDFTFDYDI 540
```

Db 653 SAACFODETIFKYS DHTMTAFGLDQSKRSASGVQEHKLRITHPSFNDFTFYDI 712
Qy 541 ALLELEKPAEYSSWVRPCLPDASHVPAGKAIWVTGHTGHTQYGGTALLIQKEIRVIN 600
Db 713 ALLELEKPAEYSSWVRPCLPDNTNHPAGKAIWVTGHTGHTGHTGALLIQKEIRVIN 772
Qy 601 OTTEENLLPOQITPRMNCVGLSGVDSGCGDGGPLSSVADGRIFQAGVWSGDCGCAQ 660
Db 773 OTTEENLLPOQITPRMNCVGLSGVDSGCGDGGPLSSVADGRIFQAGVWSGDCGCAQ 832
Qy 661 RNKPGVYTRLPFRDMIKENTGV 683
Db 833 RNKPGVYTRIPEDVDMIKEQTGV 855

RESULT 30
US-10-072-012-414
Sequence 414, Application US/10072012
Publication No. US20040033493A1
GENERAL INFORMATION:
APPLICANT: Tchenev, Velizar
APPLICANT: Spytek, Kimberly
APPLICANT: Zernhusen, Bryan
APPLICANT: Paturajan, Meera
APPLICANT: Shimkets, Richard
APPLICANT: Li, Li
APPLICANT: Gangolli, Bsha
APPLICANT: Padigaru, Murallidhara
APPLICANT: Anderson, David W.
APPLICANT: Rastelli, Luca
APPLICANT: Miller, Charles E.
APPLICANT: Teupler Jr, Raymond J.
APPLICANT: Gusev, Vladimir Y.
APPLICANT: Coleman, Steven D.
APPLICANT: Wolenc, Adam R.
APPLICANT: Pena, Carol E. A
APPLICANT: Furtak, Katarzyna
APPLICANT: Grobse, William M.
APPLICANT: Alsbrook II, John P.
APPLICANT: Lepley, Denise M.
APPLICANT: Rieger, Daniel K.
APPLICANT: Burgess, Catherine E.
TITLE OR INVENTION: Proteins and Nucleic Acids Encoding Same
FILE REFERENCE: 21402-258
CURRENT APPLICATION NUMBER: US/10/072.012
PRIOR FILING DATE: 2002-01-31
PRIOR APPLICATION NUMBER: 60/265,102
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: 60/265,514
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265,517
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265,412
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265,395
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/266,406
PRIOR FILING DATE: 2001-02-02
PRIOR APPLICATION NUMBER: 60/266,767
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: 60/267,057
PRIOR FILING DATE: 2001-02-07
PRIOR APPLICATION NUMBER: 60/266,975
PRIOR FILING DATE: 2001-02-07
PRIOR APPLICATION NUMBER: 60/267,459
PRIOR FILING DATE: 2001-02-08
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 1391
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 414
LENGTH: 855
TYPE: PRT

ORGANISM: Rattus norvegicus
US-10-072-012-414
Query Match 84.4%; Score 3197; DB 15; Length 855;
Best Local Similarity 83.0%; Pred. No. 7.8e-23;
Matches 567; Conservative 59; Mismatches 57; Indels 0; Gaps 0;
Qy 1 MAERVYMLPRPARSLKPFVTSVAPPTDQRTQRTQDSSCPGLARGVLRFTTPG 60
Db 173 MAERVYMLPRPARSLKPFVTSVAPPTDQRTQRTQDSSCPGLARGVLRFTTPG 232
Qy 61 FPDSPYPAHARCCWALRGDADSVLSLTFRSFDLASCDERSDLVTYNTLSPEPHALVQ 120
Db 233 FPNSPYPAHARCCWALRGDADSVLSLTFRSFDVAPCDGHDSDLVTVYDLSPEPHALVVR 292
Qy 121 LCGTYPPSYNLTFHSSONVLLITLITERRHGFQFATFQLPMSCCGRRLRAQGTEN 180
Db 293 LCGTFSPSYNLTFHSSONVFLVTLITNDRHGFQFATFQLPMSCCGRLRAQGTFS 352
Qy 181 SPYYPGHYPPNIDCTNIEVPPNNOHVVRPKFYLEPRRACGTCPDYVEINEKXCGE 240
Db 353 SPYYPGHYPPNIDCTNIEVPPNNOHVVRPKFYLEPRRACGTCPDYVEINEKXCGE 412
Qy 241 RSQFVYVTSNKNKLTFRSHSDQSYTDTGFLAAYLSYDSSDPGQFTCRGTGRKELRCD 300
Db 413 RSQFVYVTSNKNKLTFRSHSDQSYTDTGFLAAYLSYDSSDPGQFTCRGTGRKELRCD 472
Qy 301 GMADCTHSDDELNCSDAGHGFCTCKNFKCPLTFVVCDSVNDCCGNSDEGSCCPAQTFRC 360
Db 473 GMADCPDYSDBRHCNATHQFMCKNPFCKPLFVVCDSVNDCCGNSDEGSCCPAGSFKC 532
Qy 361 SNGKCLSKSOQCNKDCGSDGSDASCPKXNVVCTGHTYRCNLGLCLSKGNPBCDCKED 420
Db 533 SNGKCLSKSOQCNKDCGSDGSDASCDNNAVSCITYYTRCCNGCLCNKGNPBCDCKED 592
Qy 421 CSDGSDKDCDCGLRSTFRQARVVGCTDADEGEMPMQVSLHALGQHICGASLISPMVLV 480
Db 593 CSDGSDKDCDCGLRSTFRQARVVGCTDADEGEMPMQVSLHALGQHICGASLISPMVLV 652
Qy 481 SAACFYIDDRGFRYSDPTQMTAFGLHDQSORSAPGVQERKLRITHPSFNDFTFYDI 540
Db 653 SAACFODETIFKYS DHTMTAFGLDQSKRSASGVQEHKLRITHPSFNDFTFYDI 712
Qy 541 ALLELEKPAEYSSWVRPCLPDASHVPAGKAIWVTGHTGHTGHTGALLIQKEIRVIN 600
Db 713 ALLELEKPAEYSSWVRPCLPDNTNHPAGKAIWVTGHTGHTGHTGALLIQKEIRVIN 772
Qy 601 OTTEENLLPOQITPRMNCVGLSGVDSGCGDGGPLSSVADGRIFQAGVWSGDCGCAQ 660
Db 773 OTTEENLLPOQITPRMNCVGLSGVDSGCGDGGPLSSVADGRIFQAGVWSGDCGCAQ 832
Qy 661 RNKPGVYTRLPFRDMIKENTGV 683
Db 833 RNKPGVYTRIPEDVDMIKEQTGV 855

Search completed: November 29, 2004, 08:49:50
Job time : 94.8166 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using SW model

Run on: November 29, 2004, 08:20:59 : Search time 25.7568 Seconds
(without alignments)
1758.570 Million cell updates/sec

Title: US-09-936-333-5
Perfect score: 3789
Sequence: 1 MAERVYMLPPARSLKSFV.....PGVYTRLPFRDMIKENTGV 683

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database :

Issued Patents AA: *
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep: *
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep: *
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep: *
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep: *
5: /cgn2_6/ptodata/1/1aa/6C_COMB.pep: *
6: /cgn2_6/ptodata/1/1aa/backfile1.pep: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3754	99.1	855	2	US-09-027-337-2
2	3754	99.1	855	4	US-09-644-600-2
3	3754	99.1	855	4	US-09-654-600A-2
4	3123	82.4	902	4	US-09-644-600A-10
5	3123	82.4	902	4	US-09-654-600A-10
6	1319	34.8	241	4	US-09-657-986B-2
7	703.5	18.6	798	1	US-08-200-900A-2
8	703.5	18.6	798	4	US-08-794-042-2
9	703.5	18.6	798	5	PCT-US94-0616-2
10	588	15.5	407	3	US-09-734-675-4
11	560.5	14.8	492	4	US-09-685-166A-895
12	560.5	14.8	492	4	US-09-879-792-14
13	560.5	14.8	492	4	US-09-879-426-895
14	558.5	14.7	492	3	US-09-342-749-2
15	558.5	14.7	492	4	US-09-691-840-2
16	547.5	14.4	235	3	US-08-944-483-65
17	544.5	14.4	235	3	US-08-807-151-3
18	544.5	14.4	235	3	US-09-478-957-3
19	543.5	14.3	454	4	US-09-518-046-2
20	527	13.9	466	4	US-10-177-661-4
21	526.5	13.9	638	2	US-08-681-151-3
22	521.5	13.8	418	4	US-10-177-661-6
23	518.5	13.7	477	4	US-10-177-661-2
24	518.5	13.7	562	4	US-09-879-792-12
25	512.5	13.5	283	3	US-08-807-151-1
26	512.5	13.5	283	3	US-09-478-957-1
27	511.5	13.5	418	1	US-08-508-448C-25

28	511.5	13.5	418	4	US-09-370-838-82
29	511.5	13.5	418	4	US-09-370-838-83
30	511.5	13.5	418	4	US-09-854-133-82
31	511.5	13.5	418	4	US-09-854-133-83
32	510.5	13.5	418	4	US-09-370-838-62
33	510.5	13.5	418	4	US-09-854-133-62
34	505.5	13.3	376	4	US-09-820-002-2
35	504	13.3	790	4	US-08-991-761A-13
36	500.5	13.2	546	6	5200340-6
37	497.5	13.1	222	1	US-08-508-448C-19
38	496	13.1	416	2	US-09-000-846-2
39	493.5	13.0	791	1	US-08-643-219-1
40	493.5	13.0	791	3	US-08-851-350-1
41	492.5	13.0	583	4	US-09-976-594-837
42	492.5	13.0	780	1	US-08-469-486-54
43	492.5	13.0	790	2	US-08-469-658-54
44	492	13.0	417	4	US-09-820-002-4
45	491.5	13.0	791	2	US-09-131-995-1
46	491.5	13.0	791	3	US-08-832-087B-1
47	491.5	13.0	791	3	US-09-132-154-1
48	491.5	13.0	791	4	US-08-991-761A-6
49	491.5	13.0	791	4	US-08-924-287A-1
50	491.5	13.0	810	1	US-07-854-603-2
51	491.5	13.0	810	1	US-08-147-000B-29
52	491.5	13.0	810	3	US-09-086-514-1
53	491.5	13.0	810	4	US-09-192-012-5
54	491.5	13.0	810	4	US-09-403-736-1
55	491.5	13.0	810	4	US-09-701-265-1
56	491.5	13.0	810	6	5200340-8
57	491.5	13.0	812	1	US-08-248-629A-1
58	491.5	13.0	812	1	US-08-451-932-1
59	491.5	13.0	812	1	US-08-452-260-1
60	491.5	13.0	812	1	US-08-326-785-1
61	491.5	13.0	812	2	US-08-612-788-1
62	491.5	13.0	812	2	US-08-605-598B-1
63	491.5	13.0	812	2	US-08-429-743-1
64	491.5	13.0	812	2	US-08-866-735-1
65	491.5	13.0	812	3	US-09-066-028-1
66	491.5	13.0	812	4	US-09-192-012-3
67	491.5	13.0	812	4	US-09-335-325-1
68	491.5	13.0	812	4	US-08-991-761A-12
69	491.5	13.0	812	5	PCT-US95-05107-1
70	488.5	12.9	814	1	US-08-750-711-1
71	484.5	12.8	415	1	US-08-073-531B-1
72	484.5	12.8	415	2	US-08-766-288-1
73	484.5	12.8	461	3	US-08-742-877-2
74	481.5	12.7	315	4	US-09-386-653A-9
75	481.5	12.7	812	4	US-08-991-761A-7
76	481	12.7	405	3	US-09-734-675-2
77	481	12.7	810	4	US-08-991-761A-11
78	480.5	12.7	461	3	US-09-053-871A-21
79	480.5	12.7	461	4	US-10-133-907-5
80	480.5	12.7	461	6	5521070-2
81	479.5	12.7	415	1	US-08-295-411-2
82	479.5	12.7	415	5	US-08-955-471-2
83	479.5	12.7	415	5	PCT-US97-10242-2
84	478.5	12.6	256	2	US-09-027-337-3
85	478.5	12.6	256	4	US-09-644-600-3
86	478.5	12.6	256	4	US-09-654-600A-3
87	477	12.6	527	6	5520913-1
88	476.5	12.6	290	4	US-09-386-653A-7
89	476.5	12.6	328	4	US-10-140-002-222
90	476	12.6	328	4	US-09-386-642-11
91	474.5	12.5	809	4	US-08-991-761A-9
92	473.5	12.5	255	3	US-08-944-483-67
93	473.5	12.5	415	4	US-09-118-748-2
94	472	12.5	455	3	US-09-261-416-2
95	472	12.5	527	3	US-08-811-949-39
96	470	12.4	319	4	US-09-386-642-12
97	469.5	12.4	362	6	524676-5
98	469	12.4	472	2	US-08-811-949-63
99	467.5	12.3	437	2	US-08-811-949-51
100	467	12.3	527	1	US-07-609-510B-16

Sequence 82, Appl
Sequence 83, Appl
Sequence 83, Appl
Sequence 83, Appl
Sequence 62, Appl
Sequence 62, Appl
Sequence 12, Appl
Sequence 13, Appl
Sequence 19, Appl
Sequence 2, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 6, Appl
Sequence 1, Appl
Sequence 2, Appl
Sequence 2, Appl
Sequence 2, Appl
Sequence 1, Appl
Sequence 5, Appl
Sequence 54, Appl
Sequence 39, Appl
Sequence 12, Appl
Sequence No. 524676
Sequence 63, Appl
Sequence 51, Appl
Sequence 16, Appl

ALIGNMENTS

RESULT 1

US-09-027-337-2
; Sequence 2, Application US/09027337B
; Patent No. 5972616
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Tanimoto, Hirotsoshi
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease Overexpressed in
; TITLE OF INVENTION: Breat and Ovarian Carcinomas
; FILE REFERENCE: D6064
; CURRENT APPLICATION NUMBER: US/09/027,337B
; CURRENT FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 13
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Amino acid sequence of TADG-15 encoded by nucleotides
; OTHER INFORMATION: 23 to 2589 of Sequence 1
; Patent No. 5972616
US-09-027-337-2

Query Match 99.1%; Score 3754; DB 2; Length 855;

Best Local Similarity 99.3%; Pred. No. 3.9e-262; Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAERVVMLEPPRARSLSKSFVVTSVVAFPTDSKTVOQTQDNSCSFGIHLARGVELMRFTTPG 60
DB 173 MAERVVMLEPPRARSLSKSFVVTSVVAFPTDSKTVOQTQDNSCSFGIHLARGVELMRFTTPG 232
QY 61 PPDSPPYPAHARCMALRGADSVLSTFRSFDLASCDERGSDLVTYNTLSPMEPHALVQ 120
DB 233 PPDSPPYPAHARCMALRGADSVLSTFRSFDLASCDERGSDLVTYNTLSPMEPHALVQ 292
QY 121 LCGTTPPSYNTLTHSSQNVLLITLINTERRRHGFEATFFOLPRMSSCGRLRKAQGTEN 180
DB 293 LCGTTPPSYNTLTHSSQNVLLITLINTERRRHGFEATFFOLPRMSSCGRLRKAQGTEN 352
QY 181 SPYYPGHYPNIDCTWNIIEVPNNQHVAVRPFKEFYLLLEPRRACGTCPKDYVEINGEKYCGE 240
DB 353 SPYYPGHYPNIDCTWNIIEVPNNQHVAVRPFKEFYLLLEPRRACGTCPKDYVEINGEKYCGE 412
QY 241 RSQPFVVTSSNKTITVRFHSDQSYTDTGFLAELYSYSSDPCPGQFTCRGRICRKELRCD 300
DB 413 RSQPFVVTSSNKTITVRFHSDQSYTDTGFLAELYSYSSDPCPGQFTCRGRICRKELRCD 472
QY 301 GMADCTDHSDELNCSGDAHQFTCKNFKCKPLFWVCDVNDGDNDSDEGSCGPAQTFRC 360
DB 473 GMADCTDHSDELNCSGDAHQFTCKNFKCKPLFWVCDVNDGDNDSDEGSCGPAQTFRC 532
QY 361 SNGKCLSKSQOQCGKDCDGDSDDEASCPKYNVVTCTKHTRYCLNGLCLSKGNPECDGKED 420
DB 533 SNGKCLSKSQOQCGKDCDGDSDDEASCPKYNVVTCTKHTRYCLNGLCLSKGNPECDGKED 592
QY 421 CSQSDSEKDCDCLGRFTTQARVVGTDADDEGEMPWQVSLHALGQGHICGASLISPMNLV 480
DB 593 CSQSDSEKDCDCLGRFTTQARVVGTDADDEGEMPWQVSLHALGQGHICGASLISPMNLV 652
QY 481 SAHHCYIDDBGFRYSPTQMTAFGLHDOSQSAFQVQERRLRKRIISHPFNDFTFYDI 540
DB 653 SAHHCYIDDBGFRYSPTQMTAFGLHDOSQSAFQVQERRLRKRIISHPFNDFTFYDI 712
QY 541 ALLELEKPAEYSSWVRPICLPDASHVPAGKAIWVTMGHTQYGGTGALLLOKEIIRVIN 600
DB 713 ALLELEKPAEYSSWVRPICLPDASHVPAGKAIWVTMGHTQYGGTGALLLOKEIIRVIN 772
QY 601 QTTCCENLLPQOITPRMNCVGFSLGGVDSCGDSGGPLSVBADGRIFQAGVWSGDCAQ 660

RESULT 2

US-09-644-600-2
; Sequence 2, Application US/09644600
; Patent No. 6451500
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Tanimoto, Hirotsoshi
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease
; TITLE OF INVENTION: Overexpressed in Carcinomas
; FILE REFERENCE: D6064CIP/D
; CURRENT APPLICATION NUMBER: US/09/644,600
; CURRENT FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: 09/421,213
; PRIOR FILING DATE: 1999-10-20
; PRIOR APPLICATION NUMBER: 09/027,337
; PRIOR FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: TADG-15
US-09-644-600-2

Query Match 99.1%; Score 3754; DB 4; Length 855;

Best Local Similarity 99.3%; Pred. No. 3.9e-262; Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAERVVMLEPPRARSLSKSFVVTSVVAFPTDSKTVOQTQDNSCSFGIHLARGVELMRFTTPG 60
DB 173 MAERVVMLEPPRARSLSKSFVVTSVVAFPTDSKTVOQTQDNSCSFGIHLARGVELMRFTTPG 232
QY 61 PPDSPPYPAHARCMALRGADSVLSTFRSFDLASCDERGSDLVTYNTLSPMEPHALVQ 120
DB 233 PPDSPPYPAHARCMALRGADSVLSTFRSFDLASCDERGSDLVTYNTLSPMEPHALVQ 292
QY 121 LCGTTPPSYNTLTHSSQNVLLITLINTERRRHGFEATFFOLPRMSSCGRLRKAQGTEN 180
DB 293 LCGTTPPSYNTLTHSSQNVLLITLINTERRRHGFEATFFOLPRMSSCGRLRKAQGTEN 352
QY 181 SPYYPGHYPNIDCTWNIIEVPNNQHVAVRPFKEFYLLLEPRRACGTCPKDYVEINGEKYCGE 240
DB 353 SPYYPGHYPNIDCTWNIIEVPNNQHVAVRPFKEFYLLLEPRRACGTCPKDYVEINGEKYCGE 412
QY 241 RSQPFVVTSSNKTITVRFHSDQSYTDTGFLAELYSYSSDPCPGQFTCRGRICRKELRCD 300
DB 413 RSQPFVVTSSNKTITVRFHSDQSYTDTGFLAELYSYSSDPCPGQFTCRGRICRKELRCD 472
QY 301 GMADCTDHSDELNCSGDAHQFTCKNFKCKPLFWVCDVNDGDNDSDEGSCGPAQTFRC 360
DB 473 GMADCTDHSDELNCSGDAHQFTCKNFKCKPLFWVCDVNDGDNDSDEGSCGPAQTFRC 532
QY 361 SNGKCLSKSQOQCGKDCDGDSDDEASCPKYNVVTCTKHTRYCLNGLCLSKGNPECDGKED 420
DB 533 SNGKCLSKSQOQCGKDCDGDSDDEASCPKYNVVTCTKHTRYCLNGLCLSKGNPECDGKED 592
QY 421 CSQSDSEKDCDCLGRFTTQARVVGTDADDEGEMPWQVSLHALGQGHICGASLISPMNLV 480
DB 593 CSQSDSEKDCDCLGRFTTQARVVGTDADDEGEMPWQVSLHALGQGHICGASLISPMNLV 652
QY 481 SAHHCYIDDBGFRYSPTQMTAFGLHDOSQSAFQVQERRLRKRIISHPFNDFTFYDI 540
DB 653 SAHHCYIDDBGFRYSPTQMTAFGLHDOSQSAFQVQERRLRKRIISHPFNDFTFYDI 712

```
Qy 541 ALLELEKPAEYSSMWRPCLPDPASHVPAGKAIWYMGHTQYGGTALLI0KGEIRVIN 600
Db 713 ALLELEKPAEYSSMWRPCLPDPASHVPAGKAIWYMGHTQYGGTALLI0KGEIRVIN 772
Qy 601 QTTCENTLLPQOITPRMNCVGFSLSGVDSCQDSCGGLSSVEADGRIFQAGVWSMGDCAQ 660
Db 773 QTTCENTLLPQOITPRMNCVGFSLSGVDSCQDSCGGLSSVEADGRIFQAGVWSMGDCAQ 832
Qy 661 RNKPGVYTRLPPLFRDWIKENTGV 683
Db 833 RNKPGVYTRLPPLFRDWIKENTGV 855
```

RESULT 3

```
US-09-654-600A-2
; Sequence 2, Application US/09654600A
; Patent No. 6649741
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Tanimoto, Hirotsoshi
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease
; FILE REFERENCE: D6064CIP/D
; CURRENT FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 09/421,213
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 98
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: TADG-15
US-09-654-600A-2
```

```
Query Match 99.1%; Score 3754; DB 4; Length 855;
Best Local Similarity 99.3%; Pred. No. 3.9e-262;
Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
```

```
Qy 1 MAERVVMLEPPRASLSKSFVYTSVVAAPTDSKTQVORTODNSCSFGLHARGVELMRFTTGG 60
Db 173 MAERVVMLEPPRASLSKSFVYTSVVAAPTDSKTQVORTODNSCSFGLHARGVELMRFTTGG 232
Qy 61 PPDSPPYPAHARCOMALRGDADSVLSLTFRSFDLASCDERSGDLVTVYNTLSPMEPHALVQ 120
Db 233 PPDSPPYPAHARCOMALRGDADSVLSLTFRSFDLASCDERSGDLVTVYNTLSPMEPHALVQ 292
Qy 121 LCGTTPPSYNTLTFHSSQNVLLITLTNTERHHPGFATFPQLPRMSSCGGLRLKAQGTFN 180
Db 293 LCGTTPPSYNTLTFHSSQNVLLITLTNTERHHPGFATFPQLPRMSSCGGLRLKAQGTFN 352
Qy 181 SPYYPGHYPPNIDCTWNIIEVNNQHVAVRKFYLLLEPRACGTCPCPDYVEINGEKYCGE 240
Db 353 SPYYPGHYPPNIDCTWNIIEVNNQHVAVRKFYLLLEPRACGTCPCPDYVEINGEKYCGE 412
Qy 241 RSQFVYTSNSNKTIVRFHSDQSYTDGTGLAEYLSYDSSDPCPGQFTCTGRICRKEALCD 300
Db 413 RSQFVYTSNSNKTIVRFHSDQSYTDGTGLAEYLSYDSSDPCPGQFTCTGRICRKEALCD 472
Qy 301 GWADCTHSDBLNCSCDAGHOFCTCKNFKCKPLFWVCDVNDGDNSSDEGSCSPAQTRFC 360
Db 473 GWADCTHSDBLNCSCDAGHOFCTCKNFKCKPLFWVCDVNDGDNSSDEGSCSPAQTRFC 532
Qy 361 SNGKLSKSOQNGKDCGSDSDASCPKVVVCTKHTYRCLNGLCLSKNPECDGKED 420
Db 533 SNGKLSKSOQNGKDCGSDSDASCPKVVVCTKHTYRCLNGLCLSKNPECDGKED 592
Qy 421 CSDSDDEKDCGCGRLSFTROARVVGSTADAGEWQVSLHALGGHICGASLISPMNLV 480
Db 593 CSDSDDEKDCGCGRLSFTROARVVGSTADAGEWQVSLHALGGHICGASLISPMNLV 652
```

```
Qy 481 SAACHYIDBGRFPRYSDPTQMTAFGLHDOSQSRSPAGVQERLRKRIISHPFENFTPDYDI 540
Db 653 SAACHYIDBGRFPRYSDPTQMTAFGLHDOSQSRSPAGVQERLRKRIISHPFENFTPDYDI 712
Qy 541 ALLELEKPAEYSSMWRPCLPDPASHVPAGKAIWYMGHTQYGGTALLI0KGEIRVIN 600
Db 713 ALLELEKPAEYSSMWRPCLPDPASHVPAGKAIWYMGHTQYGGTALLI0KGEIRVIN 772
Qy 601 QTTCENTLLPQOITPRMNCVGFSLSGVDSCQDSCGGLSSVEADGRIFQAGVWSMGDCAQ 660
Db 773 QTTCENTLLPQOITPRMNCVGFSLSGVDSCQDSCGGLSSVEADGRIFQAGVWSMGDCAQ 832
Qy 661 RNKPGVYTRLPPLFRDWIKENTGV 683
Db 833 RNKPGVYTRLPPLFRDWIKENTGV 855
```

RESULT 4

```
US-09-644-600-10
; Sequence 10, Application US/09644600
; Patent No. 6451500
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Tanimoto, Hirotsoshi
; TITLE OF INVENTION: Overexpressed in Carcinomas
; FILE REFERENCE: D6064CIP/D
; CURRENT FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: 09/421,213
; PRIOR FILING DATE: 1999-10-20
; PRIOR APPLICATION NUMBER: 09/027,337
; PRIOR FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 98
; LENGTH: 902
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
; OTHER INFORMATION: Epitchin
US-09-644-600-10
```

```
Query Match 82.4%; Score 3123; DB 4; Length 902;
Best Local Similarity 82.9%; Pred. No. 1.1e-216;
Matches 556; Conservative 56; Mismatches 59; Indels 0; Gaps 0;
```

```
Qy 1 MAERVVMLEPPRASLSKSFVYTSVVAAPTDSKTQVORTODNSCSFGLHARGVELMRFTTGG 60
Db 173 MAERVVMLEPPRASLSKSFVYTSVVAAPTDSKTQVORTODNSCSFGLHARGVELMRFTTGG 232
Qy 61 PPDSPPYPAHARCOMALRGDADSVLSLTFRSFDLASCDERSGDLVTVYNTLSPMEPHALVQ 120
Db 233 PPDSPPYPAHARCOMALRGDADSVLSLTFRSFDLASCDERSGDLVTVYNTLSPMEPHALVQ 292
Qy 121 LCGTTPPSYNTLTFHSSQNVLLITLTNTERHHPGFATFPQLPRMSSCGGLRLKAQGTFN 180
Db 293 LCGTTPPSYNTLTFHSSQNVLLITLTNTERHHPGFATFPQLPRMSSCGGLRLKAQGTFN 352
Qy 181 SPYYPGHYPPNIDCTWNIIEVNNQHVAVRKFYLLLEPRACGTCPCPDYVEINGEKYCGE 240
Db 353 SPYYPGHYPPNIDCTWNIIEVNNQHVAVRKFYLLLEPRACGTCPCPDYVEINGEKYCGE 412
Qy 241 RSQFVYTSNSNKTIVRFHSDQSYTDGTGLAEYLSYDSSDPCPGQFTCTGRICRKEALCD 300
Db 413 RSQFVYTSNSNKTIVRFHSDQSYTDGTGLAEYLSYDSSDPCPGQFTCTGRICRKEALCD 472
Qy 301 GWADCTHSDBLNCSCDAGHOFCTCKNFKCKPLFWVCDVNDGDNSSDEGSCSPAQTRFC 360
Db 473 GWADCTHSDBLNCSCDAGHOFCTCKNFKCKPLFWVCDVNDGDNSSDEGSCSPAQTRFC 532
Qy 361 SNGKLSKSOQNGKDCGSDSDASCPKVVVCTKHTYRCLNGLCLSKNPECDGKED 420
Db 533 SNGKLSKSOQNGKDCGSDSDASCPKVVVCTKHTYRCLNGLCLSKNPECDGKED 592
```

Db 533 SNGKCLPQSOCKNGKNDKCGSDSEASCDVSVNVVSCITKYTRCQNGCLCLSKGNPECDKTD 592
Qy 421 CSDSDKDCDCGCLRSFTQARVVGTDADGEMPMQVSLHALGQGHICGASLISPMNLV 480
Db 593 CSDSDKDCDCGCLRSFTQARVVGTDADGEMPMQVSLHALGQGHICGASLISPMNLV 652
Qy 481 SAAHCYIDDRGFRYSPTQMTAFGLHDOSQASAPGVQERLKRILSHPFNDFTPDYDI 540
Db 653 SAAHCYIDDRGFRYSPTQMTAFGLHDOSQASAPGVQERLKRILSHPFNDFTPDYDI 712
Qy 541 ALLELEKPAEYSSMWRPCLPDASHVPAGKAIWVTGHTQYGTGALLIÖKGEIRVIN 600
Db 713 ALLELEKPAEYSSMWRPCLPDASHVPAGKAIWVTGHTQYGTGALLIÖKGEIRVIN 772
Qy 601 QTTCEMLPQOITPRMNCVGLSGVDSCQSDSGGLSVYADGRIFQAGVSWGDCAQ 660
Db 773 QTTCEMLPQOITPRMNCVGLSGVDSCQSDSGGLSVYADGRIFQAGVSWGDCAQ 832
Qy 661 RNKPGVYTRLP 671
Db 833 RNKPGVYTRLP 843

RESULT 5
US-09-654-600A-10
; Sequence 10, Application US/09654600A
; Patent No. 6649741
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Tamimoto, Hirotsoshi
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease
; FILE REFERENCE: D604CIP/D
; CURRENT APPLICATION NUMBER: US/09/654,600A
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 09/421,213
; PRIOR FILING DATE: 1999-10-20
; 1998-02-20
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 10
; LENGTH: 902
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
; OTHER INFORMATION: Epithin
US-09-654-600A-10

Query Match 82.4%; Score 3123; DB 4; Length 902;
Best Local Similarity 82.9%; Pred. No. 1,1e-216;
Matches 556; Conservative 56; Mismatches 59; Indels 0; Gaps 0;

Qy 1 MAERVVMLPPRASLSFVVTSVAPPTDSKYQORQDSCSGHLAGVYELMRFTTPG 60
Db 173 MAERVVMLPPRASLSFVVTSVAPPTDSKYQORQDSCSGHLAGVYELMRFTTPG 232
Qy 61 FPDSPYAHARCOVALRGDADSVLSTFRSEFDLASCEBGSGLDLYVNTLSPMEPHALVQ 120
Db 233 FPDSPYAHARCOVALRGDADSVLSTFRSEFDLASCEBGSGLDLYVNTLSPMEPHALV 292
Qy 121 LCGTPSPYNLTFRHSONVLLITLITNTERRHGFEPALPQPMSSCGGLRKAQGTEN 180
Db 293 LCGTPSPYNLTFRHSONVLLITLITNTERRHGFEPALPQPMSSCGGLRKAQGTEN 352
Qy 181 SPYRPGHYPRINDCTNMIIEVPPNNGHYVVRKFFFLILBPRAACGCPDQVYEINEKXCGE 240
Db 353 SPYRPGHYPRINDCTNMIIEVPPNNGHYVVRKFFFLILBPRAACGCPDQVYEINEKXCGE 412
Qy 241 RSQFVTVSNNSKTIYHFRHSDSYDTGTFLAEYLSDSSDDPCPGQFTCRGTGCIKELRCD 300
Db 413 RSQFVTVSNNSKTIYHFRHSDSYDTGTFLAEYLSDSSDDPCPGQFTCRGTGCIKELRCD 472
Qy 301 GMADCTDHSDELNCSDAQHQTCKNRCXPLFWVCDSDVNDGDNDSDEGCSCPAQTFRC 360

Db 473 GMADCEPDYSDERYCQGNATQFTCKNQFCXPLFWVCDSDVNDGDNDSDEGCSCPAQTFRC 532
Qy 361 SNGKCLSKQOCCNGKXDCDGDSDSEASCPKYNVYTCRKHYYRCNLNGCLSKGNPECDKTD 420
Db 533 SNGKCLPQSOCKNGKNDKCGSDSEASCDVSVNVVSCITKYTRCQNGCLCLSKGNPECDKTD 592
Qy 421 CSDSDKDCDCGCLRSFTQARVVGTDADGEMPMQVSLHALGQGHICGASLISPMNLV 480
Db 593 CSDSDKDCDCGCLRSFTQARVVGTDADGEMPMQVSLHALGQGHICGASLISPMNLV 652
Qy 481 SAAHCYIDDRGFRYSPTQMTAFGLHDOSQASAPGVQERLKRILSHPFNDFTPDYDI 540
Db 653 SAAHCYIDDRGFRYSPTQMTAFGLHDOSQASAPGVQERLKRILSHPFNDFTPDYDI 712
Qy 541 ALLELEKPAEYSSMWRPCLPDASHVPAGKAIWVTGHTQYGTGALLIÖKGEIRVIN 600
Db 713 ALLELEKPAEYSSMWRPCLPDASHVPAGKAIWVTGHTQYGTGALLIÖKGEIRVIN 772
Qy 601 QTTCEMLPQOITPRMNCVGLSGVDSCQSDSGGLSVYADGRIFQAGVSWGDCAQ 660
Db 773 QTTCEMLPQOITPRMNCVGLSGVDSCQSDSGGLSVYADGRIFQAGVSWGDCAQ 832
Qy 661 RNKPGVYTRLP 671
Db 833 RNKPGVYTRLP 843

RESULT 6
US-09-657-986B-2
; Sequence 2, Application US/09657986B
; Patent No. 6797504
; GENERAL INFORMATION:
; APPLICANT: Madison, Edwin L.
; APPLICANT: Semple, Joseph Edward
; APPLICANT: Coombs, Gary Samuel
; APPLICANT: Reiner, John Eugene
; APPLICANT: Ong, Edgar O.
; APPLICANT: Arai, Gian Luca
; TITLE OF INVENTION: Inhibitors of Serine Protease Activity of Matrilysin or
; FILE REFERENCE: Corvax 255/049
; CURRENT APPLICATION NUMBER: US/09/657,986B
; CURRENT FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 241
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-09-657-986B-2

Query Match 34.8%; Score 1319; DB 4; Length 241;
Best Local Similarity 100.0%; Pred. No. 1.8e-87;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 443 VVGSTADDEGEMPMQVSLHALGQGHICGASLISPMNLVSAHCYIDDRGFRYSPTQMTA 502
Db 1 VVGSTADDEGEMPMQVSLHALGQGHICGASLISPMNLVSAHCYIDDRGFRYSPTQMTA 60
Qy 503 FLGLHDOSQASAPGVQERLKRILSHPFNDFTPDYDIALLELEKPAEYSSMWRPCLPD 562
Db 61 FLGLHDOSQASAPGVQERLKRILSHPFNDFTPDYDIALLELEKPAEYSSMWRPCLPD 120
Qy 563 ASHVFPAGKAIWVTGHTQYGTGALLIÖKGEIRVINQTTCEMLPQOITPRMNCVGL 622
Db 121 ASHVFPAGKAIWVTGHTQYGTGALLIÖKGEIRVINQTTCEMLPQOITPRMNCVGL 180
Qy 623 SGGVDCSQSDSGGLSVYADGRIFQAGVSWGDCAQRNKPGVYTRLPFRDMIKENTG 682
Db 181 SGGVDCSQSDSGGLSVYADGRIFQAGVSWGDCAQRNKPGVYTRLPFRDMIKENTG 240
Qy 683 V 683

Db 241 V 241

RESULT 7

US-08-200-900A-2

Sequence 2, Application US/08200900A

Patent No. 5655566

GENERAL INFORMATION:

APPLICANT:

TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE

NUMBER OF SEQUENCES: 38

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Genetics Institute, Inc. - Legal Affairs

STREET: 87 Cambridgepark Drive

CITY: Cambridge

STATE: MA

COUNTRY: USA

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/200,900A

FILING DATE: 23-FEB-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Meiner, Maureen C.

REGISTRATION NUMBER: 31,544

REFERENCE/DOCKET NUMBER: GI 5201-FWC

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 876-1170 X8574

TELEFAX: (617) 876-5851

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 798 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-200-900A-2

Query Match 18.6%; Score 703.5; DB 1; Length 798;

Best Local Similarity 32.6%; Pred. No. 1.5e-42;

Matches 181; Conservative 91; Mismatches 187; Indels 97; Gaps 24;

Qy 161 QLRPMSSCGG--RIRKAGTFSNPPYPGHYPPNIDCTWNIEVNNOHVKVRFKFFYLEP 218
Db 298 ELP--TDCGGPHDLMEBNTTFTSINFPNSYPNOAFCIWNLNAQKKNIQLFQEFDELN- 354
Qy 219 RRACGTCPKDYVEI--NGEKYCGERSQFV-----VTSNKNITVRFHSDQSYTDT 266
Db 355 -----IADVEIRDE--GDSLFVAVTGPVNDVSTTNRMTVLFITDNMLAKQ 404
Qy 267 GFLLAEVYS--YDSSDPC--PGQFTCTGRCIRKELRCDGWADCTDHSDELNC-----SC 316
Db 405 GFKANFTTGYGLGPECKEDNFQCKDGEICPLVNLGCGFPHCKDGSDEAHCVRLFNGT 464
Qy 317 DAGH--GPTCKNKKCKPLFWVCDSVNDGSDSDGSCSPAGOTFRCSNGKLSLSSQCG 374
Db 465 DSSGLVQRRIS-----IMHV-----ACAEN-----WTTQISDVC-----QLLG 499
Qy 375 KDQDGDGSDASAC-----PRVNVVCTKHTYRCINGLCLSKGNPECDKEDCSGSDSEK 428
Db 500 ---LGTGNSSVPTSTGGGPFVNLNTAP-----NGSLILTPSQGC-----LDSILIL 544
Qy 429 DCD---CGLRFTQOA--RVVCGTDADGEMPMVOYSLHALQGHICGASLSIPMWLSAA 483
Db 545 QCNKYSCKKLVLTQVSEPKIVGSDSREGAMPVVALYFDQ--QVCGASLVSRLMLVSA 603
Qy 484 HCYIDDBGFRSDPTQWTAFLGLHDQGRSAPGVGERLKRILISHPFNDPTFDYDAL 543

Db 604 HCYVG-----NMESKKAIVLGLHMASNLTSPOLETBLIQIYINPHYNNRKRKNIDAM 659

Qy 544 ELEKPAEYSMAVRPICLPDASHVFPACKAIWVGHTYGGGALLIQGEIRVINQTT 603

Db 660 HLEKRVYTDIYIPICLPEENOVFPFGRICISINGWGLIYQGSTADVLQADAVPLLSNEK 719

Qy 604 CENLLPQ--QITPRMVCVGLSGVDSCQSGSPLSVEADGRIFQGVVSWGDCQARN 662

Db 720 CQQQMFENYITENNVCAGYERAGVDSCQDSGGLMCGE--NRMWLAGVTSFGYCALPN 778

Qy 663 KPGVYTRLPLEFDMIX 678

Db 779 RPYVYARVPRTFEMIQ 794

RESULT 8

US-08-794-042-2

Sequence 2, Application US/08794042

Patent No. 6746859

GENERAL INFORMATION:

APPLICANT:

TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE

NUMBER OF SEQUENCES: 38

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Genetics Institute, Inc. - Legal Affairs

STREET: 87 Cambridgepark Drive

CITY: Cambridge

STATE: MA

COUNTRY: USA

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/794,042

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/200,900

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Meiner, Maureen C.

REGISTRATION NUMBER: 31,544

REFERENCE/DOCKET NUMBER: GI 5201-FWC

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 876-1170 X8574

TELEFAX: (617) 876-5851

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 798 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-794-042-2

Query Match 18.6%; Score 703.5; DB 4; Length 798;

Best Local Similarity 32.6%; Pred. No. 1.5e-42;

Matches 181; Conservative 91; Mismatches 187; Indels 97; Gaps 24;

Qy 161 QLRPMSSCGG--RIRKAGTFSNPPYPGHYPPNIDCTWNIEVNNOHVKVRFKFFYLEP 218
Db 298 ELP--TDCGGPHDLMEBNTTFTSINFPNSYPNOAFCIWNLNAQKKNIQLFQEFDELN- 354
Qy 219 RRACGTCPKDYVEI--NGEKYCGERSQFV-----VTSNKNITVRFHSDQSYTDT 266
Db 355 -----IADVEIRDE--GDSLFVAVTGPVNDVSTTNRMTVLFITDNMLAKQ 404
Qy 267 GFLLAEVYS--YDSSDPC--PGQFTCTGRCIRKELRCDGWADCTDHSDELNC-----SC 316
Db 405 GFKANFTTGYGLGPECKEDNFQCKDGEICPLVNLGCGFPHCKDGSDEAHCVRLFNGT 464
Qy 429 DCD---CGLRFTQOA--RVVCGTDADGEMPMVOYSLHALQGHICGASLSIPMWLSAA 483
Db 545 QCNKYSCKKLVLTQVSEPKIVGSDSREGAMPVVALYFDQ--QVCGASLVSRLMLVSA 603
Qy 484 HCYIDDBGFRSDPTQWTAFLGLHDQGRSAPGVGERLKRILISHPFNDPTFDYDAL 543

Db	403	GFANPTTGVGJGIBPCKEDNFOCKDBECIPVNLCDGPFHCKDGSDEAHCVTLFNGTT	464
Qy	317	DAGH--OFTCKNKECKPFLWCDVNDGDNDSDEOGCSPAOTFRCSNGKCLSKSQONG	374
Db	465	DSSGLVQFRJQS-----IMHV-----ACAEN-----WTQISDPVC-----QLLG	499
Qy	375	KDDCGDGSDBASC-----PKVNVVTCFHTYRCLNGCLSKGNECDCKEDSGSDSK	428
Db	500	---LGTGNSSVPFTSTGGGPVYNLTAP-----NSLLITPSQOC-----LEDSLILL	544
Qy	429	DCD---CGLRSPTRQ--RVVGTDADDEGEWQVSLHALQGHICGASLISPMVLVSA	483
Db	545	QCNVYKCGKKLVYQEVSPKIVGSSDRBEGAMPWVALVFDQ--QVCGSLVSRLVLSAA	603
Qy	484	HCTIYDRGRYSPTQWTFLEGLHDQSQBSAPGVQERLKIISHPFNDFTFYDIAL	543
Db	604	HCYVG---RNNEPSKMKVGLIHMASNLTPQIETRLIDIVINPHYNNRKKRKNNDIAM	659
Qy	544	ELEKPAEYSMWVPICLPDASHVFPAGKAIWYTGWHGYQGTGALLJOKKEIVINQTT	603
Db	660	HLEMKNYTDYIOPICLPBENOVFPFGRCISLAGGALLYGGSTADVYQEDADVLSNEK	719
Qy	604	CENILPQ--QITPRMNCVGLSGGVDSQCGDSGGLPSVEADGRIFOAGVWSWEDCAQRN	662
Db	720	COQOMPEYNTENMWCAGYEAGGVDSQCGDSGGLMCGE--NNRWLLAGVTSFGVGCALPN	778
Qy	663	KRGVYTRLLPLFRDMIK	678
Db	779	RPGVYARVPRFTWIO	794

RESULT 9
 PCT-US94-00616-2
 Sequence 2, Application PC/TUS9400616
 GENERAL INFORMATION:
 APPLICANT:
 TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE
 NUMBER OF SEQUENCES: 33
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US94/00616
 FILING DATE:
 CLASSIFICATION:
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 798 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 PCT-US94-00616-2

Query Match	18.6%;	Score 703.5;	DB 5;	length 796;
Best Local Similarity	32.6%;	Pred. No. 1.5e-42;		
Matches 181; Conservative	91;	Mismatches 187;	Indels 97;	Gaps 24

[illegible][illegible]

```

RESULT 10
US-09-734-675-4
; Sequence 4, Application US/09734675
; Patent No. 6365391
; GENERAL INFORMATION:
; APPLICANT: WEBSTER, Marion et al
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CLO000862
; CURRENT APPLICATION NUMBER: US/09/734,675
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 407
; TYPE: PRT
; ORGANISM: Human
US-09-734-675-4

```

Query Match	15.5%;	Score 588;	DB 3;	Length 407;
Best Local Similarly	45.9%;	Pred. No. 1.4e-34;		
Matches 119; Conservative	38;	Mismatches 82;	Indels 20;	Gaps 6;

```

QY      432  CGLR----SFTSOARVAVGGDDADDEGEMPMQVSLHALGQGGHI CGASLISPMVLVSAACFYID  488
QY      162  CGTRSKSLTGGSIRIVGGTEVEEGEMPMQASIQMGNG-SHRCGATILINAMVLVSAACFYI-  219
Db
QY      489  DRGRFSYDPTQWTAFLGLADQSQRSAQVQERRLKRIITISHPEFNDFTFDYIALLELEXP  548
Db      220  ----TYKPAWTAWSAFGV-----TIKPKMKRGLRRIIVHEKXKPHSHDYDSLAEISSP  270
QY      549  AELYSMTVRPLCLPDASHVFPAGKAIWYGMWGHYQYGGTGALLILOGEIRVINQTCENIL  608
Db      271  VPTINAVIRVCLPDSAYEFQPEDVWFVTFGALIKNDGYSQNHLRQAVQLIADITTCNE--  328
QY      609  PQ-----QITPRMCMCVPLSGGVDSQCGDSGGLSVSEADGRIFQAGVWSWGDCAQRNPK  664
Db      329  PQAYNDATITPRMLTACGSLGKTKDCAQGDPSGCGPLVSDARDIWLAYIGVWSGDECAKPNPK  388
QY      665  GVTYTRLPFLFRDWIKENTGV  683
Db      389  GVYTRVTALRDMITSKTGI  407

```

```
RESULT 11
US-09-685-166A-895
; Sequence 895, Application US/09685166A
; Patent No. 6630305
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yugu
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stoik, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C21
; CURRENT FILING DATE: 2000-10-10
; CURRENT APPLICATION NUMBER: US/09/685.166A
; NUMBER OF SEQ ID NOS: 898
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 895
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-685-166A-895

Query Match          14.8%; Score 560.5; DB 4; Length 492;
Best Local Similarity 33.9%; Pred. No. 1.7e-32;
Matches 131; Conservative 58; Mismatches 128; Indels 69; Gaps 14;

QY 350 GCSCPAOTFRG-SNGKLSKQOCNGKDDCGDSDGDEASCPRV---NVV----- 393
DB 110 GSKCSNGIECDSSGTCINPNCDCGVSHCPGEBENRCVRLYGNPFILOMYSSQKSMH 169
QY 394 -----TCT-----KHTYRCING-----LCISKNPCECKEDKEDSD 423
DB 170 PVCDDMNENYGRACRDMGYKNNFYSSQGIIVDSGSTSPMKLNTSAGNVDIYKLYHSD 229
QY 424 GSDEK-----DCDCGLRSFTRQARVVGTDADGEMPMQVSLHALGQGHICGASLIS 475
DB 230 ACSKAVAVSLRCLACGVNLNS-SRQSRIVGGSALPGAMPQVSLH-VQVHVCGGSIIT 287
QY 476 PNMVLSAAHCYIDRGFRYSDDPTOWTAFGLHDOS-ORSAPGVQERLKRILISHPFENDF 534
DB 288 PEMVITVAHCVEKP---LNNPMHTAPAGILRQSFMEYAGYQ---VQKVISHPNYSK 340
QY 535 TFDVDIALLEKRAEYSWVRPCLPDASHVFPAGKAIWVTGNGHTQYGGTGLIIOKG 594
DB 341 TKNNDIALMKLOKPLTFNDLVKPVCLPMPGMLDPEOLCWSGATEEKGKTSSEVLNAA 400
QY 595 EIRVINTTCEN--LLPQITPRMVCVFLSGVDSCCGDSGGPLSSVEADGRI-FOAGV 651
DB 401 KVLIIETORCNSRYVYDNLITPAMICAGFLQGNVDSCCGDSGGPL--VTSNNNIWMLIGD 458
QY 652 VSWGDCGAQRNKPQVYTRLPFLFRDWI 677
DB 459 TSMGSCAKAYRPGVYGNVWFTDWI 484

RESULT 12
US-09-879-792-14
; Sequence 14, Application US/09879792
; Patent No. 6734006
; GENERAL INFORMATION:
```

```
APPLICANT: Xiao, Yonghong
APPLICANT: Gedrich, Richard
; TITLE OF INVENTION: Regulation of Human Transmembrane Sertine
; FILE REFERENCE: 02973.00035
; CURRENT APPLICATION NUMBER: US/09/879.792
; CURRENT FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/211,224
; PRIOR FILING DATE: 2000-06-13
; PRIOR APPLICATION NUMBER: US 60/283,353
; PRIOR FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: US 60/283,648
; PRIOR FILING DATE: 2001-04-16
; PRIOR APPLICATION NUMBER: PCT
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-879-792-14

Query Match          14.8%; Score 560.5; DB 4; Length 492;
Best Local Similarity 33.9%; Pred. No. 1.7e-32;
Matches 131; Conservative 58; Mismatches 128; Indels 69; Gaps 14;

QY 350 GCSCPAOTFRG-SNGKLSKQOCNGKDDCGDSDGDEASCPRV---NVV----- 393
DB 110 GSKCSNGIECDSSGTCINPNCDCGVSHCPGEBENRCVRLYGNPFILOMYSSQKSMH 169
QY 394 -----TCT-----KHTYRCING-----LCISKNPCECKEDKEDSD 423
DB 170 PVCDDMNENYGRACRDMGYKNNFYSSQGIIVDSGSTSPMKLNTSAGNVDIYKLYHSD 229
QY 424 GSDEK-----DCDCGLRSFTRQARVVGTDADGEMPMQVSLHALGQGHICGASLIS 475
DB 230 ACSKAVAVSLRCLACGVNLNS-SRQSRIVGGSALPGAMPQVSLH-VQVHVCGGSIIT 287
QY 476 PNMVLSAAHCYIDRGFRYSDDPTOWTAFGLHDOS-ORSAPGVQERLKRILISHPFENDF 534
DB 288 PEMVITVAHCVEKP---LNNPMHTAPAGILRQSFMEYAGYQ---VQKVISHPNYSK 340
QY 535 TFDVDIALLEKRAEYSWVRPCLPDASHVFPAGKAIWVTGNGHTQYGGTGLIIOKG 594
DB 341 TKNNDIALMKLOKPLTFNDLVKPVCLPMPGMLDPEOLCWSGATEEKGKTSSEVLNAA 400
QY 595 EIRVINTTCEN--LLPQITPRMVCVFLSGVDSCCGDSGGPLSSVEADGRI-FOAGV 651
DB 401 KVLIIETORCNSRYVYDNLITPAMICAGFLQGNVDSCCGDSGGPL--VTSNNNIWMLIGD 458
QY 652 VSWGDCGAQRNKPQVYTRLPFLFRDWI 677
DB 459 TSMGSCAKAYRPGVYGNVWFTDWI 484

RESULT 13
US-09-679-426-895
; Sequence 895, Application US/09679426
; Patent No. 6759515
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yugu
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stoik, John A.
; APPLICANT: Vedvick, Thomas S.
```

```

; APPLICANT: Carter, Derrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C20
; CURRENT APPLICATION NUMBER: US/09/679.426
; CURRENT FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 895
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 895
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-679-426-895

Query Match      14.8%; Score 560.5; DB 4; Length 492;
Best Local Similarity 33.9%; Pred. No. 1.7e-32;
Matches 131; Conservative 58; Mismatches 128; Indels 69; Gaps 14;

QY 350 GCSCPAQTFRG-SNGKCLSKSQCGKDGSDGSDASCPKV---NVV----- 393
DB 110 GSKCSNGIGIECDSSGTCINPSNMCDSVSHCPGSDENRCVRLYGPNTLLQMTSSQKRSWH 169
QY 394 -----TCT---KHTYRCING-----LCLSKGNPECDGKEDCSD 423
DB 170 PVCQDDMNENYGRAACRDMGYKNNFYSSQGIYDSDSGTSFMKLTNSAGNVDIYKKLYHSD 229
QY 424 GSDEK-----DCDGLRSFTTQARVVGCTDADGEGMPQVSLHALGQGHICGASLIS 475
DB 230 ACSSKAVSLRCLACGVNLNS-SRQSRIVGGSALPGAMPQVSLH-VQNVHVCGSIIIT 287
QY 476 PNMVLSAHCYIDRGRFYSPTQMTAFGLHDOS-QRSAPGVQERLRKRIISHPFNDF 534
DB 288 PEMIVTAHHCVEKP---LNNPMTAFAGILRSQSFMYGAGYQ---VQKISHPNYDSK 340
QY 535 TFDYDIALLELEKPAEYSSMVRPCLPDASHVPAGKAIWVTGHTQYGGTALILQKG 594
DB 341 TKNNDIAMKQLQKPLFNDLVKVCPLPNPQMLQPEQLCWLISGATBEKGTSEVLNAA 400
QY 595 EIRVINGTTCEN--LLPQGITPRMCMCVGLSGGVDSQCGDSGGPLSSVEADGRI-FOAGV 651
DB 401 KVLIIETQRCSNRYVDNLITTPAMICAGFLQGVNDSQCGDSGGPL--VTSNNNIWMLIGD 458
QY 652 VSMGDCGAQRKRPQVYTRLPFRDWI 677
DB 459 TSMGSGCAKAYRPGVYGNVWFDTWI 484

RESULT 14
US-09-342-749-2
; Sequence 2, Application US/09342749
; Patent No. 6166194
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Teng, David H.-F.
; APPLICANT: Myriad Genetics, Inc.
; TITLE OF INVENTION: TMRPS2 is a Tumor Suppressor
; FILE REFERENCE: 2318-202
; CURRENT APPLICATION NUMBER: US/09/342.749
; CURRENT FILING DATE: 1999-06-29
; EARLIER APPLICATION NUMBER: US 60/091,044
; EARLIER FILING DATE: 1998-06-29
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-342-749-2
```

```

Query Match      14.7%; Score 558.5; DB 3; Length 492;
Best Local Similarity 33.9%; Pred. No. 2.3e-32;
Matches 131; Conservative 57; Mismatches 129; Indels 69; Gaps 14;

QY 350 GCSCPAQTFRG-SNGKCLSKSQCGKDGSDGSDASCPKV---NVV----- 393
DB 110 GSKCSNGIGIECDSSGTCINPSNMCDSVSHCPGSDENRCVRLYGPNTLLQMTSSQKRSWH 169
QY 394 -----TCT---KHTYRCING-----LCLSKGNPECDGKEDCSD 423
DB 170 PVCQDDMNENYGRAACRDMGYKNNFYSSQGIYDSDSGTSFMKLTNSAGNVDIYKKLYHSD 229
QY 424 GSDEK-----DCDGLRSFTTQARVVGCTDADGEGMPQVSLHALGQGHICGASLIS 475
DB 230 ACSSKAVSLRCLACGVNLNS-SRQSRIVGGSALPGAMPQVSLH-VQNVHVCGSIIIT 287
QY 476 PNMVLSAHCYIDRGRFYSPTQMTAFGLHDOS-QRSAPGVQERLRKRIISHPFNDF 534
DB 288 PEMIVTAHHCVEKP---LNNPMTAFAGILRSQSFMYGAGYQ---VQKISHPNYDSK 340
QY 535 TFDYDIALLELEKPAEYSSMVRPCLPDASHVPAGKAIWVTGHTQYGGTALILQKG 594
DB 341 TKNNDIAMKQLQKPLFNDLVKVCPLPNPQMLQPEQLCWLISGATBEKGTSEVLNAA 400
QY 595 EIRVINGTTCEN--LLPQGITPRMCMCVGLSGGVDSQCGDSGGPLSSVEADGRI-FOAGV 651
DB 401 KVLIIETQRCSNRYVDNLITTPAMICAGFLQGVNDSQCGDSGGPL--VTSNNNIWMLIGD 458
QY 652 VSMGDCGAQRKRPQVYTRLPFRDWI 677
DB 459 TSMGSGCAKAYRPGVYGNVWFDTWI 484

RESULT 15
US-09-691-840-2
; Sequence 2, Application US/09691840
; Patent No. 6444419
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Teng, David H.-F.
; APPLICANT: Myriad Genetics, Inc.
; TITLE OF INVENTION: TMRPS2 is a Tumor Suppressor
; FILE REFERENCE: 2318-202
; CURRENT APPLICATION NUMBER: US/09/691,840
; CURRENT FILING DATE: 2000-10-18
; PRIOR APPLICATION NUMBER: US/09/342,749
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 60/091,044
; PRIOR FILING DATE: 1998-06-29
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-691-840-2

Query Match      14.7%; Score 558.5; DB 4; Length 492;
Best Local Similarity 33.9%; Pred. No. 2.3e-32;
Matches 131; Conservative 57; Mismatches 129; Indels 69; Gaps 14;

QY 350 GCSCPAQTFRG-SNGKCLSKSQCGKDGSDGSDASCPKV---NVV----- 393
DB 110 GSKCSNGIGIECDSSGTCINPSNMCDSVSHCPGSDENRCVRLYGPNTLLQMTSSQKRSWH 169
QY 394 -----TCT---KHTYRCING-----LCLSKGNPECDGKEDCSD 423
DB 170 PVCQDDMNENYGRAACRDMGYKNNFYSSQGIYDSDSGTSFMKLTNSAGNVDIYKKLYHSD 229
QY 424 GSDEK-----DCDGLRSFTTQARVVGCTDADGEGMPQVSLHALGQGHICGASLIS 475
```

\vdots

Db 1 IVGSDSREGAMPVVALYFDDQ-QVCASLVSRLDMLVSAHCYVG---NNMPSKKA 55
Qy 503 FLGHDSORSAPGVOERRLKRIISHPFNDFDYDIALLEKPAVSSMVPICLPD 562
Db 56 VLGHMAANSNLTSPOIFERLLIQIYINPHYNNRRKNDIAMHLEMKNYTYIPICLPE 115
Qy 563 ASHVPAKAIWVTGWHGTGYGTGALLQKEIRVINOCTTCENLLPQ-QITPRMVCVF 621
Db 116 ENQVFPPGRICSIAGMGALIIYGSTADVLQADAVPLLSNEKCOQOMPEYNTENWVCAGY 175
Qy 622 LSGGVDSQQDSSGGLPSSVADGRIFQAGVSWMGDCAQNRKRGVYTRLPFRDMIK 678
Db 176 EAGGVDSQQDSSGGLPMLCQF-NNRWLLAGVTSFGYQCALPBRGVYARVPRFTWIO 231
RESULT 18
US-09-478-957-3
; Sequence 3, Application US/09478957
; Patent No. 6350448
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; TITLE OF INVENTION: NOVEL HUMAN PROSTATE-ASSOCIATED
; TITLE OF INVENTION: PROTEASE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESS: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION NUMBER: US/09/478,957
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/807,151
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PR-0227 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 235 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: Genbank
; CLONE: 416132
; US-09-478-957-3
Query Match 14.4%; Score 544.5; DB 3; Length 235;
Best Local Similarity 43.5%; Pred. No. 1e-31;
Matches 103; Conservative 50; Mismatches 77; Indels 7; Gaps 4;

Qy 563 ASHVPAKAIWVTGWHGTGYGTGALLQKEIRVINOCTTCENLLPQ-QITPRMVCVF 621
Db 116 ENQVFPPGRICSIAGMGALIIYGSTADVLQADAVPLLSNEKCOQOMPEYNTENWVCAGY 175
Qy 622 LSGGVDSQQDSSGGLPSSVADGRIFQAGVSWMGDCAQNRKRGVYTRLPFRDMIK 678
Db 176 EAGGVDSQQDSSGGLPMLCQF-NNRWLLAGVTSFGYQCALPBRGVYARVPRFTWIO 231
RESULT 19
US-09-518-046-2
; Sequence 2, Application US/09518046
; Patent No. 6294663
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Underwood, Lowell J.
; TITLE OF INVENTION: Transmembrane Serine Protease Overexpressed
; TITLE OF INVENTION: In Ovarian Carcinoma and Uses Thereof
; FILE REFERENCE: D6192CIP
; CURRENT APPLICATION NUMBER: US/09/518,046
; CURRENT FILING DATE: 2000-03-02
; EARLIER APPLICATION NUMBER: 09/261,416
; EARLIER FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 153
; SEQ ID NO 2
; LENGTH: 454
; TYPE: PRT
; ORGANISM: Homo sapiens
; OTHER INFORMATION: complete amino acid sequence of TADG-12
; OTHER INFORMATION: protein
; US-09-518-046-2
Query Match 14.3%; Score 543.5; DB 3; Length 454;
Best Local Similarity 33.6%; Pred. No. 2.5e-31;
Matches 128; Conservative 54; Mismatches 128; Indels 71; Gaps 14;
Qy 358 FRC-SNGKCLSKSQCGKXGKDCGSDDEASCPRY-----NVVTCRKTYRC----- 402
Db 78 YRCRSFKEIETLRCDGVSDCKGDEBYCVAVGGQNAVLYFTASWTKMCSDDWKGH 137
Qy 403 -LNGCLSKGNP-----ECDG-----KEDCSD 423
Db 138 YANVACAGQLGFPSPVSSDNLVSSLEGQFREFFVSIDHLLPDDKVTALHHSVYVRGCA 197
Qy 424 GS-DEKQCD-CGRSFTROARVYGTADADGEMPMOVSILHAGHICGASLISPMVLVS 481
Db 198 GHVYTLQCTACGHR-GRSSRIYGNMSSLSPMPWASLQFGQY-HLCGGSVITPLMIT 255
Qy 482 AAHCYIDDRGFRYSDPTOWTAFLGLHDSORSAPGVOERRLKRIISHPFNDFDYDIA 541
Db 256 AAHCYVD-----LYLPKSWTIQGLVSLDNPAP---SHLVKTIYVHSKYKPKRLGNDIA 307
Qy 542 LLELEKPAEYSSWVRPCLPDASHVFPAGKAIWVTGWHGTGYGTGALLQKEIRVINO 601
Db 308 LMKIAGPLTFENEMIQVCLPENSENPPDGKVCWTSGGAEDEGDSAPVLNHAAPLISN 367
Qy 602 TTC--ENLLPQOITPRMVCVFLSGGVDSQQDSSGGLPSSVADGRIFQ-AGVSWMGDC 658
Db 368 KICNHRDVYGGTISPSMLCAGVLTGVDSCQDSSGGLP--VCEERLMLKLVGATSGIGC 425
Qy 659 AQRNKGVYTRLPFRDMIK 679
Db 426 AEVNRKGVYTRVTSFLDMIE 446
RESULT 20
US-10-177-661-4
; Sequence 4, Application US/10177661
; Patent No. 6794173
; GENERAL INFORMATION:
; APPLICANT: Anderson, Dirk M.

APPLICANT: Virca, G. Duke
TITLE OF INVENTION: DENDRITIC CELL TRANSMEMBRANE SERINE PROTEASE
FILE REFERENCE: 3256-A
CURRENT APPLICATION NUMBER: US/10/177,661
CURRENT FILING DATE: 2002-06-20
PRIOR APPLICATION NUMBER: US 60/299,606
PRIOR FILING DATE: 2001-06-20
NUMBER OF SEQ ID NOS: 6
SOFTWARE: PatentIn version 3.1
SEQ ID NO 4
LENGTH: 446
TYPE: PRT
ORGANISM: Homo sapiens
US-10-177-661-4

Query Match 13.9%; Score 527; DB 4; Length 446;
Best Local Similarity 34.5%; Pred. No. 3.8e-30;
Matches 129; Conservative 46; Mismatches 132; Indels 67; Gaps 14;

QY 333 FMVCDSDNDGSDNDEGGCGCPAOTFRCSNCKLSKSGCGKXDCGDSGDEASCPK--- 389
DB 98 FM---OGHTGIRYKEQRESCPKHAVR-----CDGVVDCKLKSDCLCVRPDW 141
QY 390 ----VNVVTCIKHTYRCINGLCLSKGNPECDKEDCDSDGDEKDC-----DCGL 434
DB 142 DKSLKTIKYSQSHQWL---PICSSNMNDSEYSEKTCQQLGFRSRSCPSGRYISLQCHGL 198
QY 435 RSFTROARVVGCTDADGEGMPQVSLHALGCGHICGASLISPNMLVSAHACYIDR--- 490
DB 199 RAMT--RRIYAGALASDSKMPQVSLH-FGTHICGGLLDAQVLTAAHCFEYTRKVL 255
QY 491 -GFRYSDPTQWTAFLG---LHDSORSAIPGVGERLRKRISHPFNFTPTPDYDIALELE 546
DB 256 EG-----WKVYAGSNLHQLPEAS-----LAEITINSYTBEDDYDIALMKLS 300
QY 547 KPAEYSSMVRPICLPDASHVPAGKAIWVTGHT-QYGTGALILQKGEIRVINTTCE 605
DB 301 KPLTSLAHHPACLPMPHQFTSLNETCMITGFKTRERDDTSFPLAEVQNLIDFKKN 360
QY 606 NLL--PQGITPRMVCVGLSGGVDSGCGSGLISVYADGRITQAGVWSKGDCAQRNK 663
DB 361 DYLVYDSYLTFRMWCAGDLRGGRDSCGDSGGL-VEQNNRWYLAGVTSWGTGCGQRNK 419
QY 664 PGVYTRLPFRDWI 677
DB 420 PGVYTKVTEVLPWI 433

RESULT 21
US-08-681-151-3
Sequence 3, Application US/08681151
Patent No. 5869637
GENERAL INFORMATION:
APPLICANT: Au-Young, Janice
APPLICANT: Bandman, Olga
APPLICANT: Braxton, Scott Michael
APPLICANT: Goll, Surya
TITLE OF INVENTION: A NOVEL HUMAN KALLIKREIN
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/681,151

FILING DATE: Herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0074US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 638 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
IMMEDIATE SOURCE:
LIBRARY: GENBANK
CLONE: 205011
US-08-681-151-3

Query Match 13.9%; Score 526.5; DB 2; Length 638;
Best Local Similarity 27.4%; Pred. No. 6.3e-30;
Matches 184; Conservative 85; Mismatches 232; Indels 171; Gaps 32;

QY 99 RGSGLVTVVNTLSMEFHVALVOLCGTYPPEYNLTFSSQNVLLITLTNTERHGF--E 156
DB 32 RGGDLAIV---TPDAQHC-QKMTFHRCLLFSF-----LAVSPKEDDKRGCFMKE 81
QY 157 ATFFQLPRMSSCG---GLRKAQGTNSPYYPGHPY-----NIDCTWNI-----V 200
DB 82 SINGTLERIRHTGASISGSHSKCGHQLSACHODIYEGIDMRGSFNISKYDSIECOKLC 141
QY 201 PNNQHV---VRFEPFLLEPRRAC-----GT-----C 225
DB 142 TNNHCOFFYYATAFRRPBYRKSCLKRSSGTPYISIKVDNLVSGFSLKSLSETGC 201
QY 226 PKD-----VYEINGEKYCGERS---QVVTSSNKITVFRHSDQSYTDGFLAEVLSYD 276
DB 202 PMDFQHFARADLNVSQVTPDAFVCRKTVCTFHPNCLFFFTYNEMWTESQBRVCLKTS 261
QY 277 SSD-PCP-----GQFTCRTRG---CIRKELRCGMDACTDHSDELNSCDAGH- 320
DB 262 KSGRPSBPPIQENAVSGVSLFTCKRARPBPCHFKI-----YSGVAFGEELNATFVGAD 316
QY 321 --QFTCKNRCKPLFWVCDSDVNDGSDNDEGGCGCPAOTFRCSNCKLSKSGCGKXDC 378
DB 317 ACDETCTKTRCOFFYSLIPDC---KAEGCKC---SLRLST----- 353
QY 379 GDGSDASCPKVVNVVTCIKHTYRCINGLCLSKGNPECDKEDCDSDGDEKDC-----DC 432
DB 354 -DGGP-----TRITY-----EAGS---SSGYSALCKVYESSDC 383
QY 433 GLNSFTROARVVGCTDADGEGMPQVSLHA--LGCGHICGASLISPNMLVSAHACYIDR 490
DB 384 TTKI---NARIVGTGINSLEGMQVSLQVKLVSQNNHMGCSIIIGRWIITAAHCF---D 437
QY 491 GFRYSDPTQWTAFLGLHDSO--RSAPGVGERLRKRISHPFNFTPTPDYDIALELEXP 548
DB 438 GIPY--PDVWRITGGIINLSIETNKP--FSSIKELIHOKTMSGSGYDIALIKQTP 492
QY 549 AEYSSMVRPICLPDASHVPAGKAIWVTGHTQYGTGALILQKGEIRVINTTCEML 608
DB 493 LNTPEQKPICLPASKADTNTIYNCWITGKYGKEREFTQNIQKATIPVPRBEQCKY 552
QY 609 PQO-ITPRMVCVGLSGGVDSGCGSGLISVYADGRITQAGVWSKGDCAQRNKGVY 667
DB 553 RDVYITKQMICAGYKEGIDACKGDSGGL-VKHSGRMQLVGTISWGECCARKEQGVY 611

QY 668 TRLEPFDWKE 679
Db 612 TKVAEYIDWLE 623

RESULT 22

US-10-177-661-6
; Sequence 6, Application US/10177661
; Patent No. 6794173
; GENERAL INFORMATION:
; APPLICANT: Anderson, Dick M.
; APPLICANT: Virca, G. Duke
; TITLE OF INVENTION: DENDRITIC CELL TRANSMEMBRANE SERINE PROTEASE
; FILE REFERENCE: 3256-A
; CURRENT APPLICATION NUMBER: US/10/177,661
; CURRENT FILING DATE: 2002-06-20
; PRIOR APPLICATION NUMBER: US 60/299,606
; PRIOR FILING DATE: 2001-06-20
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 418
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Consensus sequence
; NAME/KEY: MISC FEATURE
; LOCATION: (1)-(418)
; OTHER INFORMATION: Xaa = unknown
US-10-177-661-6

Query Match 13.8%; Score 521.5; DB 4; Length 418;
Best Local Similarity 30.9%; Pred. No. 8.9e-30;
Matches 129; Conservative 43; Mismatches 160; Indels 85; Gaps 13;

QY 329 CKPLFWCDSDVNDGSDSCDSC-----SCPQTFRCSSNGKCLSK-----SQCCNGKDDCGD 380
Db 2 CIPKXMCDDGVDCPKXGEDEKXCVXXXXXGXXXXVRLYGDXLQLQVYSSSSXXXXPVCS 61
QY 381 GSDEA-----SCPKNV-----VT 394
Db 62 NMNBSYXKXCKQWXXSAATYXSEKXXXXXGANSFKKLVSPNNLXXDXITYSXIOXL 121
QY 395 CTIKTYRCLNGCLCLSKNPECDGKEDCDSDGDEKDCDGLRSFTROARVVGTTADDEGEW 454
Db 122 XRSSXXCPSGXVSL-----QCS-----XQDCVRLNAXXMTSRITVGXXXXXGXW 168
QY 455 PMQVSLHALGQGHICGASLISPMVLVSAHCYIDDRGFRTSDP-TQMTAFGLGHDGQRS 513
Db 169 PMQVSLQXXXXGVHLCCGSLIXPKVAVLTAHCVXG---RXXKPLGXKVFAGILTXSLHX 224
QY 514 APQVQERRLKRIISHPFNDFTF-----DYDIALLEKPAEYSSMVRPCLPDAHV 567
Db 225 XPKAXKXKXVEKIIHHPYXXXXXXXKXNDIALMLKSLPFLTXDYIQLPCLPQXL 284
QY 568 PAKKAIWVTGHT-QYGGTALILQGEIRVINQTTCE--LLPQOITPRMVCVGLSG 624
Db 285 XPGTTCWIXGWAIXKXGKTSPLQDAVPLIDNKKXSYVVDXKITPRMIGAGYLEG 344
QY 625 GVDSCQDSDSGPL---SSVEADGRIPOAGVVSXWDCG-AQNRKGVYTRLLPFDWI 677
Db 345 GVDSCQDSDSGPLVCEXXXXXQNNRMWLLXGXTSWGCGXAKANRKGVTYTXFLXWI 401

RESULT 23

US-10-177-661-2
; Sequence 2, Application US/10177661
; Patent No. 6794173
; GENERAL INFORMATION:
; APPLICANT: Anderson, Dick M.
; APPLICANT: Virca, G. Duke
; TITLE OF INVENTION: DENDRITIC CELL TRANSMEMBRANE SERINE PROTEASE

; FILE REFERENCE: 3256-A
; CURRENT APPLICATION NUMBER: US/10/177,661
; CURRENT FILING DATE: 2002-06-20
; PRIOR APPLICATION NUMBER: US 60/299,606
; PRIOR FILING DATE: 2001-06-20
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 477
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-177-661-2

Query Match 13.7%; Score 518.5; DB 4; Length 477;
Best Local Similarity 32.5%; Pred. No. 1.7e-29;
Matches 135; Conservative 50; Mismatches 137; Indels 93; Gaps 17;

QY 292 CIRKELRCDGMADCTSHSDINC-----SCDAGHOTCKNKKFKPLFW--VCD 337
Db 114 CPGHAAVRCDDGVDCRKLKSDLGCVRFMDKSLIKIYSSSHQ-----WLPIC 161
QY 338 SVNDGSDSDSCDSCPAQTFRCSSNGKCLSKSQCCNGKDDCGSDASCPKVVVYCTK 397
Db 162 S--NMNDYSSEKTC-----QQLGFESAHTTIVAHHDFAHNSISLR 200
QY 398 HTYRCLNGCLCLSKNPECDGKE---DCSDGDEKDCDGLRSFTROARVVGTTADDEGE 453
Db 201 YNSTIGESLHRS---ECPGQRYISLQCS-----HGLRAMT--GRIVGALASDSK 246
QY 454 WPMQVSLHALGQGHICGASLISPMVLVSAHCYIDR-----GFRSDPTQMTAFI 505
Db 247 WPMQVSLH-FGTTHICGTLIDAQWVLTAAHCFVTRKYLEG-----MKVAVGTSN 297
QY 506 LHPQSORSAAGVQERRLKRIISHPFNDFTFDYDIALLEKPAEYSSMVRPCLPDAHV 565
Db 298 LHQLEPAAS-----IABIINSNTDEEDVDIALMLRSKVLTSAHIHPCPLMHQ 350
QY 566 VPAPKAIWVTGHT-QYGGTALILQGEIRVINQTTCE--LLPQOITPRMVCVGL 622
Db 351 TFSINETCWITGFKRETDDTKSPFLREVQVVLIFPKCNDVLYVDSYLTTPRMCAQDL 410
QY 623 SGVDSCQDSDSGPLSSVEADGRIPOAGVVSXWDCGCAONRKPVYTRLLPFDWI 677
Db 411 RGRDSCQDSDSGPL-VCEQNNRMWLLAGVTSWGTGCGQRNKPVGVTYKVTLELPMI 464

RESULT 24

US-09-879-792-12
; Sequence 12, Application US/09879792
; Patent No. 6734006
; GENERAL INFORMATION:
; APPLICANT: Xiao, Yonghong
; APPLICANT: Gedrich, Richard
; TITLE OF INVENTION: Regulation of Human Transmembrane Serine
; FILE REFERENCE: 02973.00035
; CURRENT APPLICATION NUMBER: US/09/879,792
; CURRENT FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/211,224
; PRIOR FILING DATE: 2000-06-13
; PRIOR APPLICATION NUMBER: US 60/283,353
; PRIOR FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: US 60/283,648
; PRIOR FILING DATE: 2001-04-16
; PRIOR APPLICATION NUMBER: PCT
; PRIOR FILING DATE: 2001-06-12
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 562
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-879-792-12 (Docket No. 6734006 LIO-81-WO)

STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: SCOR0701
CLONE: 556016
US-09-478-957-1

Query Match 13.5%; Score 512.5; DB 3; Length 283;
Best Local Similarity 39.5%; Pred. No. 2.5e-29;
Matches 111; Conservative 48; Mismatches 99; Indels 23; Gaps 9;

QY 409 SKNPEDCGKEDCDSDGDE-----KDCDGLRSTFRQARVVGTDADGEGMPQVSL 460
DB 6 SAGVVDYKLYHSDACSSRAVVALRCIACGVNINS-SROSRIYGGESALPGAMPQVSL 64
QY 461 HALGQGHICGASLISPMVLVSAAHCIYIDRGFRYSPTQMTAFGLHDSQSRAPVOE 519
DB 65 H-VQNVAVCCGSITTPETVTAHCVCKP---LNPMHTAFGLIRQSGMFGAGYO- 118
QY 520 RLKRIISHPEFNFTEDYDIALLEKPAEYSMWVPICLPDASHVFPAGKAIWVGW 579
DB 119 --VEKVISHPYDSEKTNADIALMKLQKPLFNDLVKPVCLPNQGMLOPEQLCMISGW 176
QY 580 HTQYGGTALLQGEIRVINQTTCCN--LIPQITPRMVCVGLSGVDSQSGSGPL 637
DB 177 ATEKRGITSEVLNAKAVLLETORCNSRYVDNLITPAMICAGFLQGVNDSQSGSGXL 236
QY 638 SSVEADGR-IFQAGVSMGDCGCAQRNKGYYTRLPLFRDNI 677
DB 237 --VTSKNIMWLIDTSMGSCAKAYRPGYIGNWETDNI 275

RESULT 27
US-08-508-448C-25
Sequence 25, Application US/08508448C
Patent No. 5804410

GENERAL INFORMATION:
APPLICANT: Kazuyoshi YAMAOKA et al.
TITLE OF INVENTION: NUCLEIC ACID SEQUENCE ENCODING
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wenderoth, Lind & Ponack
STREET: 805 Fifteenth Street, N.W., #700
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/508.448C
FILING DATE: July 28, 1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-8850
TELEFAX:

INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 418 amino acids
TYPE: amino acid
STRANDEDNESS: single

TOPOLOGY: linear
US-08-508-448C-25

Query Match 13.5%; Score 511.5; DB 1; Length 418;
Best Local Similarity 37.4%; Pred. No. 4.6e-29;
Matches 108; Conservative 52; Mismatches 100; Indels 29; Gaps 10;

QY 409 SKNPEDCGKEDCDSDGDE-----KDCDGLRSTFRQARVVGTDADGEGMPQVSL 460
DB 145 NSGNLFINPSTETISLTDQAANWMLNECGAGPDLTTSBQRILGSTEABEGSWPQVSL 204
QY 461 HALGQGHICGASLISPMVLVSAAHCIYIDRGFRYSPTQMTAFGLHDSQSRAPVOE 520
DB 205 R-LNNAHHCAGSLINNMWILTAHCFRSN-----SNPRMIATSGI-----STFPKLR-M 253
QY 521 RLKRIISHPEFNFTEDYDIALLEKPAEYSMWVPICLPDASHVFPAGKAIWVGW 580
DB 254 RVNRIILHNNYSATHEINDIALVRLSNSVTFTDHSVCLPAATQNI PGSTAYVTGKGA 313
QY 581 TOYGGTALLQGEIRVINQTTCCNLPQO---ITPRMVCVGLSGVDSQSGSGP 636
DB 314 QEYAGHTVPELRQGVRIISNDVCN--APHSYNGALISGMLCAGVPQGVDAQSGSGGP 371
QY 637 LSSVEADGR-IFQAGVSMGDCGCAQRNKGYYTRLPLFRDIKENTGV 683
DB 372 L--VQEDSRILMPTIVGIVSWDQGLPDRKGYVTRVATYLDWIRQDTGI 418

RESULT 28
US-09-370-838-82
Sequence 82, Application US/09370838
Patent No. 6444425

GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Mohamath, Roach
APPLICANT: Secrist, Heather
TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF
FILE REFERENCE: 210121.475C1
CURRENT APPLICATION NUMBER: US/09/370.838
CURRENT FILING DATE: 1999-08-09
EARLIER APPLICATION NUMBER: US 09/285,323
EARLIER FILING DATE: 1999-04-02
NUMBER OF SEQ ID NOS: 289
SOFTWARE: PaacSeq for Windows Version 3.0
SEQ ID NO 82
LENGTH: 418
TYPE: PRT
ORGANISM: Homo sapien
US-09-370-838-82

Query Match 13.5%; Score 511.5; DB 4; Length 418;
Best Local Similarity 37.4%; Pred. No. 4.6e-29;
Matches 108; Conservative 52; Mismatches 100; Indels 29; Gaps 10;

QY 409 SKNPEDCGKEDCDSDGDE-----KDCDGLRSTFRQARVVGTDADGEGMPQVSL 460
DB 145 NSGNLFINPSTETISLTDQAANWMLNECGAGPDLTTSBQRILGSTEABEGSWPQVSL 204
QY 461 HALGQGHICGASLISPMVLVSAAHCIYIDRGFRYSPTQMTAFGLHDSQSRAPVOE 520
DB 205 R-LNNAHHCAGSLINNMWILTAHCFRSN-----SNPRMIATSGI-----STFPKLR-M 253
QY 521 RLKRIISHPEFNFTEDYDIALLEKPAEYSMWVPICLPDASHVFPAGKAIWVGW 580
DB 254 RVNRIILHNNYSATHEINDIALVRLSNSVTFTDHSVCLPAATQNI PGSTAYVTGKGA 313
QY 581 TOYGGTALLQGEIRVINQTTCCNLPQO---ITPRMVCVGLSGVDSQSGSGP 636
DB 314 QEYAGHTVPELRQGVRIISNDVCN--APHSYNGALISGMLCAGVPQGVDAQSGSGGP 371
QY 637 LSSVEADGR-IFQAGVSMGDCGCAQRNKGYYTRLPLFRDIKENTGV 683

Db 372 L-VQEDSRRLMFIIVGIVSWGDCGLPDKPGVYTRVATYLDWIRQGTGI 418

RESULT 29

US-09-370-838-83

Sequence 83, Application US/09370838

Patent No. 6444425

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.

APPLICANT: Lodes, Michael J.

APPLICANT: Mohamath, Radoch

APPLICANT: Secrist, Heather

TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF

TITLE OF INVENTION: LUNG CANCER AND METHODS FOR THEIR USE

FILE REFERENCE: 210121.475C1

CURRENT APPLICATION NUMBER: US/09/370.838

CURRENT FILING DATE: 1999-08-09

EARLIER APPLICATION NUMBER: US 09/285.323

NUMBER OF SEQ ID NOS: 289

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 83

LENGTH: 418

TYPE: PRT

ORGANISM: Homo sapien

US-09-370-838-83

Query Match 13.5%; Score 511.5; DB 4; Length 418;
Best Local Similarity 37.4%; Pred. No. 4.6e-29;
Matches 108; Conservative 52; Mismatches 100; Indels 29; Gaps 10;

Qy 409 SKGNECDGKEDSGDSE-----KDCDGLRSFT-ROARVVGCTDADGEMPMQVSL 460
Db 145 NSGNLEINPSTEITSLTDQAAANWLINECAGPDLITLSEQRILGTEAEBSMPQVSL 204
Qy 461 HALGQGHICGASLISPMVLVSAHCYIDRGFRYSDFPTQWTAFLGLHDQSORAPGVQER 520
Db 205 R-LNNAHCGSLINNMWILTAHCFRSN-----SNPRDWIATSGI-----STTFPKLR-M 253
Qy 521 RLKRIISHPPFNDFPYDIALLEKPAEYSSMWRICLPDASHVFPAGKAIWYTGWGH 580
Db 254 RVRNILLNNYKSATHEINDIALVRLNLSVTFKDIHVCCLPAATONIIPGSTAYVTGWA 313
Qy 581 TOYGTGALLIOKEIRVINTTCENLPPQ-----ITPRMVCVGLSGVDSGCGDSGCP 636
Db 314 QEYAGHTVPELRQGVRIISNDVCN--APHSYNGAILSGMLCAGVPQGVADACGDSGCP 371
Qy 637 LSSVEADGR--IFQAGVSWGDCGCAQRNKPQVYTRLPFRDWIKENTGV 683
Db 372 L-VQEDSRRLMFIIVGIVSWGDCGLPDKPGVYTRVATYLDWIRQGTGI 418

RESULT 30

US-09-854-133-82

Sequence 82, Application US/09854133

Patent No. 6759508

GENERAL INFORMATION:

APPLICANT: Lodes, Michael J.

APPLICANT: Mohamath, Radoch

APPLICANT: Henderson, Robert A.

APPLICANT: Benson, Darin R.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR

TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER

FILE REFERENCE: 210121.475C10

CURRENT APPLICATION NUMBER: US/09/854.133

NUMBER OF SEQ ID NOS: 735

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 82

LENGTH: 418

TYPE: PRT

ORGANISM: Homo sapien

US-09-854-133-82

Query Match 13.5%; Score 511.5; DB 4; Length 418;
Best Local Similarity 37.4%; Pred. No. 4.6e-29;
Matches 108; Conservative 52; Mismatches 100; Indels 29; Gaps 10;

Qy 409 SKGNECDGKEDSGDSE-----KDCDGLRSFT-ROARVVGCTDADGEMPMQVSL 460
Db 145 NSGNLEINPSTEITSLTDQAAANWLINECAGPDLITLSEQRILGTEAEBSMPQVSL 204
Qy 461 HALGQGHICGASLISPMVLVSAHCYIDRGFRYSDFPTQWTAFLGLHDQSORAPGVQER 520
Db 205 R-LNNAHCGSLINNMWILTAHCFRSN-----SNPRDWIATSGI-----STTFPKLR-M 253
Qy 521 RLKRIISHPPFNDFPYDIALLEKPAEYSSMWRICLPDASHVFPAGKAIWYTGWGH 580
Db 254 RVRNILLNNYKSATHEINDIALVRLNLSVTFKDIHVCCLPAATONIIPGSTAYVTGWA 313
Qy 581 TOYGTGALLIOKEIRVINTTCENLPPQ-----ITPRMVCVGLSGVDSGCGDSGCP 636
Db 314 QEYAGHTVPELRQGVRIISNDVCN--APHSYNGAILSGMLCAGVPQGVADACGDSGCP 371
Qy 637 LSSVEADGR--IFQAGVSWGDCGCAQRNKPQVYTRLPFRDWIKENTGV 683
Db 372 L-VQEDSRRLMFIIVGIVSWGDCGLPDKPGVYTRVATYLDWIRQGTGI 418

Search completed: November 29, 2004, 08:34:26
Job time : 30.7568 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 29, 2004, 08:14:53 / Search time 98.1424 Seconds

(without alignments)
2496.496 Million cell updates/sec

Title: US-09-936-333-5

Perfect score: 3789

Sequence: 1 MAERVMVLPARRASLSKFV.....PGVYTRLPFRDKENTGV 683

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729239 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first-100 summaries

Database :

A_Geneseq_23Sep04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3760	99.2	683	3 AAB19551	AAB19551 Human mat
2	3760	99.2	762	3 AAY90284	AAY90284 Human pep
3	3760	99.2	855	3 AAB19552	AAB19552 Human mat
4	3760	99.2	855	4 AAB35465	AAB35465 Human mem
5	3760	99.2	855	5 ADI16817	ADI16817 Human NOV
6	3760	99.2	855	5 ADI16813	ADI16813 Human NOV
7	3760	99.2	855	5 ADI16876	ADI16876 Human NOV
8	3760	99.2	855	7 ADN39867	ADN39867 Cancer/an
9	3760	99.2	855	8 ADN04754	ADN04754 Antipepti
10	3756	99.1	855	5 ADI16884	ADI16884 Human NOV
11	3756	99.1	855	5 ADI16818	ADI16818 Human NOV
12	3754	99.1	757	5 ADI16508	ADI16508 Human NOV
13	3754	99.1	855	2 AAY06671	AAY06671 Tumour an
14	3754	99.1	855	4 AAB98500	AAB98500 Human TAD
15	3754	99.1	855	4 AAE06930	AAE06930 Human mem
16	3754	99.1	855	5 AAO22929	AAO22929 Type II t
17	3754	99.1	855	5 ADI16816	ADI16816 Human NOV
18	3754	99.1	855	5 ADI16882	ADI16882 Human NOV
19	3754	99.1	855	5 ADI16875	ADI16875 Human NOV
20	3754	99.1	855	6 ABB56619	ABB56619 Human mem
21	3754	99.1	855	6 AAO30146	AAO30146 Human mem
22	3754	99.1	855	6 AAE29820	AAE29820 Human mem
23	3754	99.1	855	6 AAE29791	AAE29791 Human mem
24	3754	99.1	855	6 ABB72376	ABB72376 Transmemb
25	3754	99.1	855	7 ADB97551	ADB97551 Human MTS

26	3754	99.1	855	7 ADI10371	ADI10371 Human cel
27	3754	99.1	855	8 ADG65326	ADG65326 Human MTS
28	3754	99.1	855	8 ADI28861	ADI28861 Human mat
29	3754	99.1	855	8 ADJ46895	ADJ46895 Human tra
30	3745	98.8	851	4 AAM25628	AAM25628 Human pro
31	3745	98.8	851	4 AAB11428	AB11428 Human mem
32	3726	98.3	853	8 ADO55145	ADO55145 Protein #
33	3719	98.2	757	8 ADN42162	ADN42162 Human nov
34	3549	93.7	922	4 ABG21442	ABG21442 Novel hum
35	3259.5	86.0	782	5 ABG96427	ABG96427 Human ova
36	3209	84.7	855	5 AAE23083	AAE23083 Epithin p
37	3209	84.7	855	5 ADI16819	ADI16819 Murine NO
38	3209	84.7	855	5 ADI16877	ADI16877 Murine NO
39	3197	84.4	855	5 ADI16820	ADI16820 Rat NOVX
40	3197	84.4	855	5 ADI16881	ADI16881 Rat NOVX
41	3197	84.4	855	5 ADI16878	ADI16878 Rat NOVX
42	3123	82.4	902	4 AAB98507	AAB98507 Murine ep
43	3123	82.4	902	5 AAB080517	AAB080517 Mouse epi
44	3123	82.4	902	5 AAU77549	AAU77549 Murine ty
45	2797	73.8	620	3 AAB43748	AAB43748 Human can
46	2254.5	59.5	845	5 ADI16879	ADI16879 African c
47	1319	34.8	241	4 AAE06936	AAE06936 Human mem
48	1319	34.8	241	5 AAO22930	AAO22930 Type II t
49	1319	34.8	241	5 AAE22837	AAE22837 Human mat
50	1319	34.8	241	6 ABB56620	ABB56620 Human MTS
51	1319	34.8	241	6 AAO30147	AAO30147 Human mem
52	1319	34.8	241	6 AAE29821	AAE29821 Human mem
53	1319	34.8	241	6 AAE29792	AAE29792 Human mem
54	1319	34.8	241	6 ABB72377	ABB72377 Transmemb
55	1319	34.8	241	7 ADA25176	ADA25176 Human MTS
56	1319	34.8	241	7 ADB97553	ADB97553 Human MTS
57	1319	34.8	241	8 ADI28862	ADI28862 Human mat
58	1319	34.7	241	2 AAM22987	AAM22987 Human ser
59	1029	27.2	850	2 AAU82750	AAU82750 Amino aci
60	1026	27.1	693	7 AAE38320	AAE38320 Human mem
61	1020	26.9	648	7 AAE38322	AAE38322 Human mem
62	1019.5	26.9	706	7 AAE38321	AAE38321 Human mem
63	1019.5	26.9	843	5 AAU77552	AAU77552 Hman memb
64	1019.5	26.9	843	7 AAE38319	AAE38319 Human mem
65	1015.5	26.8	572	8 ADG31207	ADG31207 Novel mou
66	1014	26.8	799	5 ADI16874	ADI16874 Murine NO
67	1014	26.8	799	5 ADI16880	ADI16880 Murine NO
68	1007	26.6	795	5 ABB43952	ABB43952 Human PRO
69	1005	26.5	802	2 AAY41710	AAY41710 Human PRO
70	1005	26.5	802	3 AAB44266	AAB44266 Human PRO
71	1005	26.5	802	3 AAB24052	AAB24052 Human PRO
72	1005	26.5	802	5 AAB62755	AAB62755 Amino aci
73	1005	26.5	802	6 ABO25212	ABO25212 Novel hum
74	1005	26.5	802	6 ABO72218	ABO72218 Novel hum
75	1005	26.5	802	6 ABO44898	ABO44898 Human sec
76	1005	26.5	802	6 ABO61096	ABO61096 Human sec
77	1005	26.5	802	6 ABO80365	ABO80365 Human sec
78	1005	26.5	802	6 ADA24708	ADA24708 Novel hum
79	1005	26.5	802	6 ABO19667	ABO19667 Novel hum
80	1005	26.5	802	6 ADA12369	ADA12369 Human sec
81	1005	26.5	802	6 ABO19558	ABO19558 Novel hum
82	1005	26.5	802	7 ADB73675	ADB73675 Human PRO
83	1005	26.5	802	7 ADB76391	ADB76391 Human PRO
84	1005	26.5	802	7 ADC43817	ADC43817 Human sec
85	1005	26.5	802	7 ADC61577	ADC61577 Human sec
86	1005	26.5	802	7 ADC63541	ADC63541 Human sec
87	1005	26.5	802	7 ADC65641	ADC65641 Human sec
88	1005	26.5	802	7 ADC68765	ADC68765 Human sec
89	1005	26.5	802	7 ADC62825	ADC62825 Human sec
90	1005	26.5	802	7 ADC67890	ADC67890 Human sec
91	1005	26.5	802	7 ADC41210	ADC41210 Human sec
92	1005	26.5	802	7 ADC67265	ADC67265 Human sec
93	1005	26.5	802	7 ADC62201	ADC62201 Human sec
94	1005	26.5	802	7 ADC41834	ADC41834 Human sec
95	1005	26.5	802	7 ADB49203	ADB49203 Human sec
96	1005	26.5	802	7 ADB35257	ADB35257 Human sec
97	1005	26.5	802	7 ADE16371	ADE16371 Human sec
98	1005	26.5	802	7 ADD72986	ADD72986 Human sec

99 1005 26.5 802 7 ADD72344 Add72344 Human sec
100 1005 26.5 802 7 ADEL6995 ADEL6995 Human sec

ALIGNMENTS

```
RESULT 1
AAB19551
ID AAB19551 standard; protein; 683 AA.
AC AAB19551;
XX
XX 22-JAN-2001 (first entry)
XX
XX Human matrilysin (truncated form).
XX
XX Matrilysin; serine protease; human; breast cancer; pre-malignancy;
XX actinic keratosis; leukoplakia; Barrett's epithelium;
XX columnar metaplasia; ulcerative colitis; Bowenoid papulosis;
XX adenomatous colorectal polyp; Ovarian erythroplasia;
XX vulvar intraepithelial neoplasia; tumour; metastasis; therapy.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX Peptide 1..31
XX Protein /label=Signal_peptide
XX Region /label=Mature_protein
XX Peptide /note="complement subcomponent 1r and 1s (C1r/s) region"
XX Modified-site /note="RGD motif"
XX Region /note="N-glycosylated"
XX Region /note="complement subcomponent 1r and 1s (C1r/s) region"
XX Region /note="LDL-receptor type region"
XX Modified-site /note="N-glycosylated"
XX Region /note="LDL-receptor type region"
XX Region /note="LDL-receptor type region"
XX Region /note="LDL-receptor type region"
XX Region /note="LDL-receptor type region"
XX Domain /note="LDL-receptor type region"
XX Cleavage-site /label=Catalytic_domain
XX Disulfide-bond /note="conserved proteolytic activation site"
XX Active-site 484
XX Active-site /note="forms catalytic triad with Asp-539 and Ser-633"
XX Disulfide-bond /note="forms catalytic triad with His-484 and Ser-633"
XX Disulfide-bond 604..618
XX Active-site 629..658
XX /note="forms catalytic triad with His-484 and Asp-539"
XX
XX WO20005332-A1.
XX
XX 14-SEP-2000.
XX
XX 10-MAR-2000; 2000WO-US006111.
XX
XX 12-MAR-1999; 99US-0124006P.
XX
XX (GBOU ) UNIV GEORGETOWN.
XX
XX Dickson RB, Lin C, Johnson M, Wang S, Enyedy I;
```

```
XX
XX WPI; 2000-594268/56.
XX DR N-PSDB; AAB88492.
XX
XX Treating malignancies, premalignant and pathologic conditions in a
XX PT subject, comprises administering matrilysin modulating agent.
XX
XX Claim 14; Fig 9; 116p; English.
XX
XX The present sequence is that of the truncated form of human matrilysin, a
XX trypsin-like protease, as deduced from cDNA (see AAB88492) obtained from
XX human breast cancer cell cDNA by RT-PCR. The full-length form (see
XX AAB19552) has an additional 172 N-terminal amino acids. Either form can
XX be produced in transfected or transfected cells using a claimed method.
XX The zymogen (inactive) form of matrilysin is a single-chain protein. The
XX active 2-chain form strongly interacts with fragments of a Kunitz-type
XX serine protease inhibitor (hepatocyte growth factor activator inhibitor,
XX HAI-1) to form SDS-stable complexes. In breast cancer cells, matrilysin
XX is present mainly as the uncomplexed form. Only the complexed matrilysin
XX is detected in human milk. The invention is directed to a method of
XX detecting a malignancy or a pre-malignant lesion in breast or other
XX tissue by detecting the presence of single- or 2-chain forms of
XX matrilysin in the tissue. The object is to inhibit tumour onset, tumour
XX growth and metastasis. Malignancies and pre-malignant conditions
XX characterised by expression of the zymogen or activated form of
XX matrilysin are treated by administering an inhibitor of matrilysin,
XX especially a Bowman-Birk inhibitor. The pre-malignant condition is
XX atypical ductal hyperplasia of the breast, actinic keratosis,
XX leukoplakia, Barrett's epithelium of the esophagus, ulcerative colitis,
XX adenomatous colorectal polyps, erythroplasia of the Quervist, Bowen's
XX disease, Bowenoid papulosis, vulvar intraepithelial neoplasia or
XX dysplastic changes to the cervix. The invention also provides methods for
XX in vivo or in vitro diagnosis of malignancy or pre-malignant lesion, and
XX methods of identifying matrilysin modulators, including activators and
XX inhibitors
XX
XX Sequence 683 AA:
XX
XX Query Match 99.2%; Score 3760; DB 3; Length 683;
XX Best Local Similarity 99.4%; Pred. No. 3,7e-241;
XX Matches 679; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
XX
XX 1 MAERVYMLPPRRSLKSPVYVVAFPPTDSKVORTQNSCGFGLHARGVELMRTTGG 60
XX |||||||
XX 1 MAERVYMLPPRRSLKSPVYVVAFPPTDSKVORTQNSCGFGLHARGVELMRTTGG 60
XX
XX 61 FPDSPYPAHARCOMALRGADSVLSLTPRSFDLASCDEKSDLYVYNTLSPMEPALVQ 120
XX |||||||
XX 61 FPDSPYPAHARCOMALRGADSVLSLTPRSFDLASCDEKSDLYVYNTLSPMEPALVQ 120
XX
XX 121 LCGTTPPSYVLTFFHSSQVLLITLINTERRRHGFATFPQJLRMSCCGRLLKAQGTEN 180
XX |||||||
XX 121 LCGTTPPSYVLTFFHSSQVLLITLINTERRRHGFATFPQJLRMSCCGRLLKAQGTEN 180
XX
XX 181 SPYFGHYPPNIDCTNIEVPNNQHYKRPFFYLLEPPRACGTCRKYVEINTEGRCCGE 240
XX |||||||
XX 181 SPYFGHYPPNIDCTNIEVPNNQHYKRPFFYLLEPPRACGTCRKYVEINTEGRCCGE 240
XX
XX 241 RSQFVYVTSNKKITVFFHSDQSTDTGFLAEYLSYSSPPCPQFPCRTGRCTRKRLKRD 300
XX |||||||
XX 241 RSQFVYVTSNKKITVFFHSDQSTDTGFLAEYLSYSSPPCPQFPCRTGRCTRKRLKRD 300
XX
XX 301 GMADCTDHSDELNCSGDAGHQTCKNKFCKPLFWVCDVNDGDNBDEQCSGPAQTFRC 360
XX |||||||
XX 301 GMADCTDHSDELNCSGDAGHQTCKNKFCKPLFWVCDVNDGDNBDEQCSGPAQTFRC 360
XX
XX 361 SNKCTSKSQOQCKGKDCGDSDEASCPRVNVYTCRKYRCLNGCLSKNPECCGKXD 420
XX |||||||
XX 361 SNKCTSKSQOQCKGKDCGDSDEASCPRVNVYTCRKYRCLNGCLSKNPECCGKXD 420
XX
XX 421 CSDSDEKDCDGLRFTROARVVGSTDADEGEWPMQVSLHALGQGHICGASLISPMWLV 480
XX |||||||
XX 421 CSDSDEKDCDGLRFTROARVVGSTDADEGEWPMQVSLHALGQGHICGASLISPMWLV 480
XX
XX Db
```

Qy	481	SAAHGCIYIDRGGRYSPTQWTFGLGHDSQSRAPGVQSRRLKRIISHPEFNDFTPYDI	540
Dd	481	SAAHGCIYIDRGGRISDPTWTFGLGHDSQSRAPGVQSRRLKRIISHPEFNDFTPYDI	540
Qy	541	ALLELEKPAEYSSWMVRPICLPDASHVFPAGKAIWVGWGHGTGYGTGALLILQGEIRVIN	600
Dd	541	ALLELEKPAEYSSWMVRPICLPDASHVFPAGKAIWVGWGHGTGYGTGALLILQGEIRVIN	600
Qy	601	QTTGENTLPQOITTPRMVCVGLSGVDSCCGDSGGFLSSVLEADGRIFQAGVSWGDCCAQ	660
Dd	601	QTTGENTLPPQOITTPRMVCVGLSGVDSCCGDSGGFLSSVLEADGRIFQAGVSWGDCCAQ	660
Qy	661	RNKGVYTRLPFRDWIKENTGV	683
Dd	661	RNKGVYTRLPFRDWIKENTGV	683
RESULT 2			
ID	AA90284		
AC	AA90284	standard; protein; 762 AA.	
XX	AA90284;		
DT	24-OCT-2000	(first entry)	
XX			
DE		Human peptidase, HPEP-1 protein sequence.	
XX			
KW		Human; peptidase; cell proliferative disorder; arteriosclerosis;	
KW		psoriasis; myelofibrosis; cancer; autoimmune disorder; Crohn's disease;	
KW		inflammatory disorder; AIDS; anaemia; allergy; asthma; arteriosclerosis;	
KW		Grave's disease; multiple sclerosis; scleroderma; infection; diabetes;	
KM		metabolic disorder; Addison's disease; cystic fibrosis; diagnosis;	
KM		glycogen storage disease; obesity; therapy; HPEP-1.	
XX			
XX		Homo sapiens.	
OS			
XX		MO200042201-A2.	
PN			
XX		20-JUL-2000.	
PD			
XX		11-JAN-2000; 2000MO-US000641.	
PF			
XX		11-JAN-1999; 99US-0172247P.	
PR		03-MAY-1999; 99US-0132253P.	
PR		27-MAY-1999; 99US-0136653P.	
PR			
XX		(INCY-) INCYTE PHARM INC.	
XX			
PI		Bandman O, Hillman JL, Tang YT, Azimzai Y, Baughn MR, Lal P;	
P1		Pye H, Lu DAM;	
XX			
DR		WPI; 2000-482832/42.	
DR		N-PSDB; AAA37657.	
XX			
PT		An isolated polypeptide for diagnosis, prevention and treatment of cell	
PT		proliferative, autoimmune/inflammatory and metabolic disorders comprises	
PT		a sequence encoding a human peptidase.	
XX			
PS		Claim 2, Page 91-93; 131pp; English.	
XX			
CC		This sequence represents a human peptidase, designated HPEP-1. The	
CC		invention relates to 18 human peptidases designated HPEP-1 to HPEP-18,	
CC		respectively. The peptidases can be used for treating a disease or	
CC		condition associated with decreased expression or over expression of	
CC		functional human peptidases. The diseases that can be diagnosed,	
CC		prevented and treated include cell proliferative disorders (such as	
CC		arteriosclerosis, psoriasis, myelofibrosis, and cancers),	
CC		autoimmune/inflammatory disorders (such as AIDS, anaemia, allergies,	
CC		Crohn's disease, asthma, arteriosclerosis, Grave's disease, multiple	
CC		sclerosis, and scleroderma), infections, and metabolic disorders (such as	
CC		Addison's disease, diabetes, cystic fibrosis, glycogen storage diseases	
CC		and obesity)	

[illegible]

XX Key Location/Qualifiers
 FH Region 214..327
 FT Peptide /note= "complement subcomponent 1r and 1s (Clr/s) region"
 FT /note= ".251
 FT Modified-site /note= "RGD motif"
 FT /note= "302
 FT Region /note= "N-glycosylated"
 FT 340..440
 FT /note= "complement subcomponent 1r and 1s (Clr/s) region"
 FT Region 452..486
 FT /note= "LDL-receptor type region"
 FT Modified-site 485
 FT /note= "N-glycosylated"
 FT 487..523
 FT Region /note= "LDL-receptor type region"
 FT 524..559
 FT Region /note= "LDL-receptor type region"
 FT 566..602
 FT Region /note= "LDL-receptor type region"
 FT 603..855
 FT Domain /label= Catalytic_domain
 FT Cleavage-site 614..618
 FT /note= "conserved proteolytic activation site"
 FT
 PN MO200053232-A1.
 PD 14-SBP-2000.
 PF 10-MAR-2000; 2000MO-US006111.
 PR 12-MAR-1999; 99US-0124006P.
 PA (GBOU) UNIV GEORGETOWN.
 PI Dickson RB, Lin C, Johnson M, Wang S, Enyedy I;
 DR WPI; 2000-594268/56.
 DR N-PSDB; AAA88493.
 XX
 PT Treating malignancies, premalignant and pathologic conditions in a
 PT subject, comprises administering matrixase modulating agent.
 PS Claim 14; Fig 16; 116pp; English.
 XX
 CC The present sequence is that of the full-length form of human matrixase,
 CC a trypsin-like protease. This has an additional 172 amino acids compared
 CC with the truncated form of matrixase given in AAB19551. Either form can
 CC be produced in transformed or transfected cells using a claimed method.
 CC The zymogen (inactive) form of matrixase is a single-chain protein. The
 CC active 2-chain form strongly interacts with fragments of a Kunitz-type
 CC serine protease inhibitor (hepatocyte growth factor activator inhibitor,
 CC HAI-1) to form SDS-stable complexes. In breast cancer cells, matrixase
 CC is present mainly as the uncomplexed form. Only the complexed matrixase
 CC is detected in human milk. The invention is directed to a method of
 CC detecting a malignancy or a pre-malignant lesion in breast or other
 CC tissue by detecting the presence of single- or 2-chain forms of
 CC matrixase in the tissue. The object is to inhibit tumour onset, tumour
 CC growth and metastasis. Malignancies and pre-malignant conditions
 CC characterised by expression of the zymogen or activated form of
 CC matrixase are treated by administering an inhibitor of matrixase,
 CC especially a Bowman-Birk inhibitor. The pre-malignant condition is
 CC atypical ductal hyperplasia of the breast, actinic keratosis,
 CC leukoplakia, Barrett's epithelium of the esophagus, ulcerative colitis,
 CC adenomatous colorectal polyps, erythroplasia of the Queyrat, Bowen's
 CC disease, Bowenoid papulosis, vulvar intraepithelial neoplasia or
 CC dysplastic changes to the cervix. The invention also provides methods for
 CC in vivo or in vitro diagnosis of malignancy or pre-malignant lesion, and
 CC methods of identifying matrixase modulators, including activators and
 XX inhibitors
 XX Sequence 855 AA;

Query Match 99.2%; Score 3760; DB 3; Length 855;
 Best Local Similarity 99.4%; Pred. No. 4..7e-241;
 Matches 679; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 MAERVVMLEPRRSLSKSFVTVSVAFPTDSKTVORTONSCSGFGLHARGVELMRFTTPG 60
 DB 173 MAERVVMLEPRRSLSKSFVTVSVAFPTDSKTVORTONSCSGFGLHARGVELMRFTTPG 222
 QY 61 FPDSPYPAHARCCOMALRGADSVSLTFPSFDIASCDERGSIDLVTYNTLSPMEPHALVQ 120
 DB 233 FPDSPYPAHARCCOMALRGADSVSLTFPSFDIASCDERGSIDLVTYNTLSPMEPHALVQ 292
 QY 121 LCGTTPPSYNLTFHSSQNVLLITLINTERRHGFATFPQLPRMSSCGRLRKAQGTN 180
 DB 293 LCGTTPPSYNLTFHSSQNVLLITLINTERRHGFATFPQLPRMSSCGRLRKAQGTN 352
 QY 181 SPYPGHYHPNIDCTNNIEVPNNQHVKFKPFYLLPPRACGTCPKDYVEINGEKXCGE 240
 DB 353 SPYPGHYHPNIDCTNNIEVPNNQHVKFKPFYLLPPRACGTCPKDYVEINGEKXCGE 412
 QY 241 RSQFVVTSSNKTIVRFHSDSYTDTGFLAEYLSYDSDPCPGQFTCRGRCIRKELRCD 300
 DB 413 RSQFVVTSSNKTIVRFHSDSYTDTGFLAEYLSYDSDPCPGQFTCRGRCIRKELRCD 472
 QY 301 GWADCTDHSDELNCSGDAGHQFTCKNKFCKPLFWCDSYNDCCDNBDEGCSCPAQTFRC 360
 DB 473 GWADCTDHSDELNCSGDAGHQFTCKNKFCKPLFWCDSYNDCCDNBDEGCSCPAQTFRC 532
 QY 361 SNKGKLSKSGQCGKDCGSDGDEASCPRKNVYVCTKHTYRCCLNGCLSKGNDECGKED 420
 DB 533 SNKGKLSKSGQCGKDCGSDGDEASCPRKNVYVCTKHTYRCCLNGCLSKGNDECGKED 592
 QY 421 CSDGSEKDCDGLRSFTQARVVGTTDADGEMPMQVSLHALGQGHICGASLISPNMLV 480
 DB 593 CSDGSEKDCDGLRSFTQARVVGTTDADGEMPMQVSLHALGQGHICGASLISPNMLV 652
 QY 481 SAHACTIDRGFRFSPTQMTAFGLHDOSQSRAPBQVORRLKRIISHPFNDFDYDI 540
 DB 653 SAHACTIDRGFRFSPTQMTAFGLHDOSQSRAPBQVORRLKRIISHPFNDFDYDI 712
 QY 541 ALLELEKPAEYSMWVPICLPDASHVFPAGKAIWVWGHTQYGGTALILQKGRIRVIN 600
 DB 713 ALLELEKPAEYSMWVPICLPDASHVFPAGKAIWVWGHTQYGGTALILQKGRIRVIN 772
 QY 601 QTTCEMLLPQOITPRMVCVGLSGVDSCQDGSGLSVADGRIFQAGVSWDGCQAQ 660
 DB 773 QTTCEMLLPQOITPRMVCVGLSGVDSCQDGSGLSVADGRIFQAGVSWDGCQAQ 832
 QY 661 RNKRGVYTRLPFRDWIKENTGV 683
 DB 833 RNKRGVYTRLPFRDWIKENTGV 855
 RESULT 4
 AAB35465
 ID AAB35465 standard; protein; 855 AA.
 XX
 AC AAB35465;
 XX
 DT 06-JUN-2001 (first entry)
 XX
 DE Human membrane-type serine protease MT-SPL.
 KM Human; membrane-type serine protease; MT-SPL; cancer.
 XX
 OS Homo sapiens.
 XX
 PN WO200123524-A2.
 XX
 PD 05-APR-2001.
 XX
 PF 02-OCT-2000; 2000MO-US027250.
 XX

PR 30-SEP-1999; 99US-00410362.
 XX (REGC) UNIV CALIFORNIA.
 XX Craik CS, Takeuchi T, Shuman M;
 XX WPI: 2001-245002/25.
 DR N-PSDB; AAF28099.
 XX
 PT New nucleic acid encoding a membrane type serine protease, useful for the
 PT diagnosis, prognosis and treatment of cancer, particularly metastatic
 PT cancers.
 XX
 PS Claim 7, Fig 1, 102pp; English.
 XX
 CC The present invention provides the protein and coding sequences for the
 CC novel human membrane-type serine protease MT-Sp1. Increased expression of
 CC this protein is associated with cancer, and so the sequences can be used
 CC in cancer diagnosis and the identification of treatments. The present
 CC sequence is the MT-Sp1 protein
 CC
 XX Sequence 855 AA;
 SQ
 Query Match 99.2%; Score 3760; DB 4; Length 855;
 Best Local Similarity 99.4%; Pred. No. 4,7e-241;
 Matches 679; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 MAEERVVMLPPRARSLKSFVTVSVAFPTDSKTVQRTQDNCSFGLHARGVELMRFTTPG 60
 DB 173 MAEERVVMLPPRARSLKSFVTVSVAFPTDSKTVQRTQDNCSFGLHARGVELMRFTTPG 232
 QY 61 PPDSFPYAHARCOMALRGDADSVSLTPRSFDLASCDERGDLYTVYNTLSPMEPHALVQ 120
 DB 233 PPDSFPYAHARCOMALRGDADSVSLTPRSFDLASCDERGDLYTVYNTLSPMEPHALVQ 292
 QY 121 LCGYTPPSYNTLTFSSQNVLLITLITNTERRHPGFEATFOLPMNSCGGLRKAQGTFN 180
 DB 293 LCGYTPPSYNTLTFSSQNVLLITLITNTERRHPGFEATFOLPMNSCGGLRKAQGTFN 352
 QY 181 SPYYPGHYPNIDCTWNIIEVNNOHVKKRFPYLLBPRACGTCPKDYVEINSEKTCGE 240
 DB 353 SPYYPGHYPNIDCTWNIIEVNNOHVKKRFPYLLBPRACGTCPKDYVEINSEKTCGE 412
 QY 241 RSQVTVTSNSNKTIVRRHSHSDSYTDYTGFLAFLSYDSSDPFGQFTCTGTCIKELACD 300
 DB 413 RSQVTVTSNSNKTIVRRHSHSDSYTDYTGFLAFLSYDSSDPFGQFTCTGTCIKELACD 472
 QY 301 GMADCTHSDSLNCSCDAGHOFCTCKNFKCKPLFWVCDVNDGNSDQGSCEPAQTPRC 360
 DB 473 GMADCTHSDSLNCSCDAGHOFCTCKNFKCKPLFWVCDVNDGNSDQGSCEPAQTPRC 532
 QY 361 SNGKCLSKSQOQCNKGDCGSDSEASCPKVVVYCTKTKTYRCLNGLCLSKGNPECDKED 420
 DB 533 SNGKCLSKSQOQCNKGDCGSDSEASCPKVVVYCTKTKTYRCLNGLCLSKGNPECDKED 592
 QY 421 CSDSDSDERDCCGLRSFTKQARVYGGTADGCEPMQVSLHALQGHICGSLISPMNLV 480
 DB 593 CSDSDSDERDCCGLRSFTKQARVYGGTADGCEPMQVSLHALQGHICGSLISPMNLV 652
 QY 481 SAACHYIDDRGFRYSDFQTWTAFLGLHDQSORSAPGVVERLKIISHPFNDFTFDYDI 540
 DB 653 SAACHYIDDRGFRYSDFQTWTAFLGLHDQSORSAPGVVERLKIISHPFNDFTFDYDI 712
 QY 541 ALLELEKPAEYSSWVRPICLPDASHVPDAGRAIWTGNGHTQYGGTGLILQKEIRVIN 600
 DB 713 ALLELEKPAEYSSWVRPICLPDASHVPDAGRAIWTGNGHTQYGGTGLILQKEIRVIN 772
 QY 601 QTTCEMLLPQOITTRAMCVGLSGVDSCQDSSGGLSSVADGRIPAGATVSNMGDGAQ 660
 DB 773 QTTCEMLLPQOITTRAMCVGLSGVDSCQDSSGGLSSVADGRIPAGATVSNMGDGAQ 832
 QY 661 RNKPGVYTRLPLFRDWIKENTGV 683

DB 833 RNKPGVYTRLPLFRDWIKENTGV 855
 RESULT 5
 ID AD116817
 XX AD116817 standard; protein; 855 AA.
 XX
 AC AD116817;
 XX
 DT 15-APR-2004 (first entry)
 XX
 DE Human NOXV protein homologue Segid 353.
 XX
 KW human; NOXV; cardiomyopathy; atherosclerosis; cancer; diabetes;
 KW inflammation; autoimmune disorder; allergy; blood disorder;
 KW acquired immunodeficiency syndrome; AIDS; obesity; achmia;
 KW immunoglobulin (Ig)A nephropathy; cirrhosis; arthritis;
 KW Alzheimer's disease; infection; str.
 XX
 OS Homo sapiens.
 XX
 FN W0200268649-A2.
 XX
 PD 06-SEP-2002.
 XX
 PF 31-JAN-2002; 2002MO-US002785.
 XX
 XX 31-JAN-2001; 2001US-0265395P.
 PR 31-JAN-2001; 2001US-0265412P.
 PR 31-JAN-2001; 2001US-0265514P.
 PR 31-JAN-2001; 2001US-0265517P.
 PR 02-FEB-2001; 2001US-0266406P.
 PR 05-FEB-2001; 2001US-0266757P.
 PR 07-FEB-2001; 2001US-0266975P.
 PR 08-FEB-2001; 2001US-0267057P.
 PR 09-FEB-2001; 2001US-0267459P.
 PR 15-FEB-2001; 2001US-0267823P.
 PR 26-FEB-2001; 2001US-0271664P.
 PR 27-FEB-2001; 2001US-0271839P.
 PR 27-FEB-2001; 2001US-0271855P.
 PR 02-MAR-2001; 2001US-0272788P.
 PR 02-MAR-2001; 2001US-0273046P.
 PR 14-MAR-2001; 2001US-0275925P.
 PR 14-MAR-2001; 2001US-0275947P.
 PR 14-MAR-2001; 2001US-0275950P.
 PR 14-MAR-2001; 2001US-0275959P.
 PR 15-MAR-2001; 2001US-0276448P.
 PR 15-MAR-2001; 2001US-0276450P.
 PR 16-MAR-2001; 2001US-0276397P.
 PR 16-MAR-2001; 2001US-0276768P.
 PR 20-MAR-2001; 2001US-0278652P.
 PR 26-MAR-2001; 2001US-0278775P.
 PR 26-MAR-2001; 2001US-0278778P.
 PR 29-MAR-2001; 2001US-0279882P.
 PR 29-MAR-2001; 2001US-0279884P.
 PR 30-MAR-2001; 2001US-0280147P.
 PR 11-APR-2001; 2001US-0282932P.
 PR 11-APR-2001; 2001US-0283083P.
 PR 20-APR-2001; 2001US-0285133P.
 PR 23-APR-2001; 2001US-0285749P.
 PR 03-MAY-2001; 2001US-0286372P.
 PR 03-MAY-2001; 2001US-0286504P.
 PR 29-MAY-2001; 2001US-0294047P.
 PR 30-MAY-2001; 2001US-0294473P.
 PR 08-JUN-2001; 2001US-0296964P.
 PR 18-JUN-2001; 2001US-0298959P.
 PR 19-JUN-2001; 2001US-0299334P.
 PR 13-AUG-2001; 2001US-0312020P.
 PR 16-AUG-2001; 2001US-0312889P.
 PR 16-AUG-2001; 2001US-0312908P.
 PR 21-AUG-2001; 2001US-0313390P.
 PR 28-AUG-2001; 2001US-0315470P.

PR 31-AUG-2001; 2001US-0316447P.
PR 07-SEP-2001; 2001US-0318115P.
PR 07-SEP-2001; 2001US-0318118P.
PR 12-SEP-2001; 2001US-0318740P.
PR 19-SEP-2001; 2001US-0323379P.
PR 18-OCT-2001; 2001US-0320245P.
PR 18-OCT-2001; 2001US-0330308P.
PR 14-NOV-2001; 2001US-0332701P.
XX
PA (CURA-) CURAGEN CORP.
XX
XX Tchernev VT, Szytek KA, Zernhusen BD, Patturajan M, Shinketsa RA;
PI Li L, Ganggoli EA, Padigaru M, Anderson DW, Rastelli L, Miller CE;
PI Gerlach VL, Taupier RJ, Gusev VY, Colman SD, Wolenc AR, Pena CEA;
PI Futrak K, Grosse WM, Alebrock JP, Lepley DM, Rieger DK, Burgess CE;
XX
XX WPI; 2002-706998/76.
XX
XX New NOXV polypeptides and nucleic acids, useful for preventing or
PT treating NOXV-associated disorders, e.g. cancer, cardiomyopathy,
PT atherosclerosis, or diabetes, and in chromosome mapping, tissue typing or
PT pharmacogenomics.
XX
XX Disclosure; SEQ ID NO 353; 1498bp; English.
XX
XX This invention relates to a novel nucleic acids, and encoded polypeptides
CC thereof, which have properties related to the stimulation of biochemical
CC or physiological responses in a cell, tissue, organ or organism.
CC Specifically, it refers to the use of biologically active fragments for
CC diagnostic and prognostic assays and furthermore in the treatment of
CC diverse pathological conditions. The present invention describes novel
CC human and murine NOXV proteins, as well as methods to modulate their
CC expression using antisense oligos, ribozymes and peptide nucleic acids.
CC The NOXV polypeptides, polynucleotides and antibodies are useful in
CC treating or preventing NOXV-associated disorders, e.g. cardiomyopathy,
CC atherosclerosis, cancer and diabetes. Furthermore, they may be used in
CC treating or preventing diseases such as inflammation, autoimmune
CC disorders, allergies, blood disorders, acquired immunodeficiency syndrome
CC (AIDS), obesity, asthma, immunoglobulin (Ig)A nephropathy, cirrhosis,
CC arthritis, Alzheimer's disease, infections, stroke, muscular dystrophy
CC and epilepsy. Accordingly, these molecules have many activities including
CC cytotaxtic, cardiant, antiinflammatory, immunosuppressive, anti-allergic,
CC haemostatic, anti-HIV, antidiabetic, antiatherosclerotic, anorectic,
CC antiaesthetic, nephrotropic, antiaerthritic, hepatotropic,
CC neuroprotective, nootropic, antibacterial, vitruide, antiparasitic,
CC relaxant and anticonvulsant. In addition, they are useful in screening
CC assays to identify small molecules that modulate or inhibit, for example,
CC neurogenesis, wound healing and angiogenesis. The nucleic acids are also
CC used as in chromosome mapping, tissue typing, preventive medicine and
CC pharmacogenomics. This polypeptide is a homologue of a human NOXV protein
CC of the invention.
XX
XX Sequence 855 AA:
SQ
Query Match 99.2%; Score 3760; DB 5; Length 855;
Best Local Similarity 99.4%; Pred. No. 4,7e-241;
Matches 679; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 241 RSQFVVTNSNKTITVPHSDQSYTDGFLAEYLISYSSDPCPGQFTCTGRCIRKELRCD 300
DB 413 RSQFVVTNSNKTITVPHSDQSYTDGFLAEYLISYSSDPCPGQFTCTGRCIRKELRCD 472
QY 301 GMAADCTDHSDELNCSCDAGHQFTCKNFKCPLRWVCDSDVNDCCGNSDEQSCSPAQTFRC 360
DB 473 GMAADCTDHSDELNCSCDAGHQFTCKNFKCPLRWVCDSDVNDCCGNSDEQSCSPAQTFRC 532
QY 361 SNGKCLSKSQCGKDCGSDGDEASCPRKNVYVCTGHTRCANGCLSKGNECDCKED 420
DB 533 SNGKCLSKSQCGKDCGSDGDEASCPRKNVYVCTGHTRCANGCLSKGNECDCKED 592
QY 421 CSDGSEKDCDCGIRSFETRRQARVVGTDADGEMPMQVSLHALGQGHICGASLISPMVLV 480
DB 593 CSDGSEKDCDCGIRSFETRRQARVVGTDADGEMPMQVSLHALGQGHICGASLISPMVLV 652
QY 481 SAACHYIDRGFRYSPTQMTAFGLHDSORSAFQVQERLKRILSHFPNDFTPYDI 540
DB 653 SAACHYIDRGFRYSPTQMTAFGLHDSORSAFQVQERLKRILSHFPNDFTPYDI 712
QY 541 ALLELEKPAEYSMWPRICLPDASHVPAGKALWVGWGHTOXGGTGALILQGEIRVIN 600
DB 713 ALLELEKPAEYSMWPRICLPDASHVPAGKALWVGWGHTOXGGTGALILQGEIRVIN 772
QY 601 QTTCEMLPQOITPRMNCVGLSGVDSCGDSCGFLSSVEADGRIFQAGVSMGDCAQ 660
DB 773 QTTCEMLPQOITPRMNCVGLSGVDSCGDSCGFLSSVEADGRIFQAGVSMGDCAQ 832
QY 661 RNRPGVYTRPLPRDINKENTGV 683
DB 833 RNRPGVYTRPLPRDINKENTGV 855
RESULT 6
ID AD116883 standard; protein; 855 AA.
XX AD116883;
XX
DT 15-APR-2004 (first entry)
XX
XX Human NOXV protein homologue SegID 419.
XX
XX human; NOXV; cardiomyopathy; atherosclerosis; cancer; diabetes;
KW inflammation; autoimmune disorder; allergy; blood disorder;
KW acquired immunodeficiency syndrome; AIDS; obesity; asthma;
KW immunoglobulin (Ig)A nephropathy; cirrhosis; arthritis;
KW Alzheimer's disease; infection; stc.
XX
XX Homo sapiens.
XX
XX WO200268649-A2.
XX
XX 06-SEP-2002.
XX
XX 31-JAN-2002; 2002WO-US002785.
XX
PR 31-JAN-2001; 2001US-0265195P.
PR 31-JAN-2001; 2001US-0265412P.
PR 31-JAN-2001; 2001US-0265514P.
PR 31-JAN-2001; 2001US-0265517P.
PR 02-FEB-2001; 2001US-0266406P.
PR 05-FEB-2001; 2001US-0266767P.
PR 07-FEB-2001; 2001US-0266975P.
PR 07-FEB-2001; 2001US-0267057P.
PR 08-FEB-2001; 2001US-0267459P.
PR 09-FEB-2001; 2001US-0267823P.
PR 15-FEB-2001; 2001US-0268974P.
PR 26-FEB-2001; 2001US-0271664P.
PR 27-FEB-2001; 2001US-0271839P.
PR 27-FEB-2001; 2001US-0271855P.
PR 02-MAR-2001; 2001US-0272788P.

PR 02-MAR-2001; 2001US-0273046P.
 PR 14-MAR-2001; 2001US-0275925P.
 PR 14-MAR-2001; 2001US-0275947P.
 PR 14-MAR-2001; 2001US-0275950P.
 PR 14-MAR-2001; 2001US-0275989P.
 PR 15-MAR-2001; 2001US-0276448P.
 PR 15-MAR-2001; 2001US-0276450P.
 PR 16-MAR-2001; 2001US-0276397P.
 PR 16-MAR-2001; 2001US-0276768P.
 PR 20-MAR-2001; 2001US-0278652P.
 PR 26-MAR-2001; 2001US-0278775P.
 PR 29-MAR-2001; 2001US-0278882P.
 PR 29-MAR-2001; 2001US-0279884P.
 PR 30-MAR-2001; 2001US-0280147P.
 PR 11-APR-2001; 2001US-0282922P.
 PR 11-APR-2001; 2001US-0283083P.
 PR 20-APR-2001; 2001US-0285133P.
 PR 23-APR-2001; 2001US-0285749P.
 PR 03-MAY-2001; 2001US-0288327P.
 PR 03-MAY-2001; 2001US-0288504P.
 PR 29-MAY-2001; 2001US-0294047P.
 PR 30-MAY-2001; 2001US-0294473P.
 PR 08-JUN-2001; 2001US-0296964P.
 PR 18-JUN-2001; 2001US-0298959P.
 PR 19-JUN-2001; 2001US-0299324P.
 PR 13-AUG-2001; 2001US-0312020P.
 PR 16-AUG-2001; 2001US-0312889P.
 PR 16-AUG-2001; 2001US-0312908P.
 PR 21-AUG-2001; 2001US-0313390P.
 PR 28-AUG-2001; 2001US-0315470P.
 PR 31-AUG-2001; 2001US-0316447P.
 PR 07-SEP-2001; 2001US-0318115P.
 PR 07-SEP-2001; 2001US-0318118P.
 PR 12-SEP-2001; 2001US-0318740P.
 PR 19-SEP-2001; 2001US-0323379P.
 PR 18-OCT-2001; 2001US-0330245P.
 PR 18-OCT-2001; 2001US-0330308P.
 PR 14-NOV-2001; 2001US-0332701P.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Tchernev VT, Soytek KA, Zehrhusen BD, Patnuraajan M, Shinkete RA;
 PI Li L, Gangolli EA, Padigam M, Anderson DM, Rastelli L, Miller CE;
 PI Gerlach VM, Taupier RJ, Gusev VY, Colman SD, Wolenc AK, Pena CE;
 PI Futak K, Grosse WM, Alsobrook JP, Lepley DM, Rieger DK, Burgess CE;
 XX
 XX WPI; 2002-706998/76.
 DR
 XX
 PT New NOVX polypeptides and nucleic acids, useful for preventing or
 PT treating NOVX-associated disorders, e.g. cancer, cardiomyopathy,
 PT atherosclerosis, or diabetes, and in chromosome mapping, tissue typing or
 PT pharmacogenomics.
 PT
 PS Disclosure: SEQ ID NO 419; 1498bp; English.
 XX
 XX This invention relates to a novel nucleic acid, and encoded polypeptides
 CC thereof, which have properties related to the stimulation of biochemical
 CC or physiological responses in a cell, tissue, organ or organism.
 CC Specifically, it refers to the use of biologically active fragments for
 CC diagnostic and prognostic assays and furthermore in the treatment of
 CC diverse pathological conditions. The present invention describes novel
 CC human and murine NOVX proteins, as well as methods to modulate their
 CC expression using antisense oligos, ribozymes and peptide nucleic acids.
 CC The NOVX polypeptides, polynucleotides and antibodies are useful in
 CC treating or preventing NOVX-associated disorders, e.g. cardiomyopathy,
 CC atherosclerosis, cancer and diabetes. Furthermore, they may be used in
 CC treating or preventing diseases such as inflammation, autoimmune
 CC disorders, allergies, blood disorders, acquired immunodeficiency syndrome
 CC (AIDS), obesity, asthma, immunoglobulin (IgA) nephropathy, cirrhosis,
 CC arthritis, Alzheimer's disease, infections, stroke, muscular dystrophy
 CC and epilepsy. Accordingly, these molecules have many activities including
 CC cyostatic, cardiant, antiinflammatory, immunosuppressive, antiallergic,

CC haemostatic, anti-HIV, antidiabetic, antiarteriosclerotic, anorectic,
 CC antiaesthetic, nephrotoxic, antiallergic, hepatotropic,
 CC neuroprotective, nootropic, antibacterial, virucide, antiparasitic,
 CC relaxant and anticonvulsant. In addition, they are useful in screening
 CC assays to identify small molecules that modulate or inhibit, for example,
 CC neurogenesis, wound healing and angiogenesis. The nucleic acids are also
 CC used as in chromosome mapping, tissue typing, preventive medicine and
 CC pharmacogenomics. This polypeptide is a homologue of a human NOVX protein
 CC of the invention.
 XX
 SQ Sequence 855 AA;
 Query Match 99.2%; Score 3760; DB 5; Length 855;
 Best Local Similarity 99.4%; Pred. No. 4.7e-241;
 Matches 679; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 MAERVVMLEPPRARSLSFVVTSVVAPPTDSKTQRTQDSCSFGHARGVELMRFTTPG 60
 DB 173 MAERVVMLEPPRARSLSFVVTSVVAPPTDSKTQRTQDSCSFGHARGVELMRFTTPG 232
 QY 61 PPDSPYPAHARCGVALGADSVSLTFRSFDLASCBERSDLYTVNTLSPMEPHALVQ 120
 DB 233 PPDSPYPAHARCGVALGADSVSLTFRSFDLASCBERSDLYTVNTLSPMEPHALVQ 292
 QY 121 LCGTTPPSVNLTFHSSQVNLTLITNTERHHPGEATFPPLPMSSCGRLRAQGFN 180
 DB 293 LCGTTPPSVNLTFHSSQVNLTLITNTERHHPGEATFPPLPMSSCGRLRAQGFN 352
 QY 181 SPYYPGHYPNIDCTWNIIEVPNNQHVVRFEFYLLBPRRACGTCPKDYVEINEKTCGE 240
 DB 353 SPYYPGHYPNIDCTWNIIEVPNNQHVVRFEFYLLBPGVAGTCPDYVEINEKTCGE 412
 QY 241 RSQFVVTNSNKTIVRHSQSYTDYTGFLAEYLSYDSDPCPGQFTRTGCIKEKRC 300
 DB 413 RSQFVVTNSNKTIVRHSQSYTDYTGFLAEYLSYDSDPCPGQFTRTGCIKEKRC 472
 QY 301 GMADCTHSDDELNCSGAGHOFCKNFKCKPLFVVCDSVNDCCGNSDEGSCCPAOTFRC 360
 DB 473 GMADCTHSDDELNCSGAGHOFCKNFKCKPLFVVCDSVNDCCGNSDEGSCCPAOTFRC 532
 QY 361 SNGKLSKSQCCNGKDCGDSDEASCPKVVVVTCTHRTYCLNGLCISKNPEDCKED 420
 DB 533 SNGKLSKSQCCNGKDCGDSDEASCPKVVVVTCTHRTYCLNGLCISKNPEDCKED 592
 QY 421 CSDSDSDKDCDCLRSTTRQARVVGCTDADGEMPMOVSILHALGCGHICGASLISPMVLV 480
 DB 593 CSDSDSDKDCDCLRSTTRQARVVGCTDADGEMPMOVSILHALGCGHICGASLISPMVLV 652
 QY 481 SAACHTIDDRGFRSDPTOWTAPLGLHDOSORSAFGVBRRLKRIISHPFNDTFPYDI 540
 DB 653 SAACHTIDDRGFRSDPTOWTAPLGLHDOSORSAFGVBRRLKRIISHPFNDTFPYDI 712
 QY 541 ALLELEPAEYSSVWRPICLPDASHVFPAGKAIWVTMGHTOYGGTALLLOKEIRVIN 600
 DB 713 ALLELEPAEYSSVWRPICLPDASHVFPAGKAIWVTMGHTOYGGTALLLOKEIRVIN 772
 QY 601 QTTCENTLLPQITPRMNCVGFSLGCVDSGCGSGLPSVADGRITQAGVSWGDCAQ 660
 DB 773 QTTCENTLLPQITPRMNCVGFSLGCVDSGCGSGLPSVADGRITQAGVSWGDCAQ 832
 QY 661 RNKRGVYTRLPFLFDWIKENTGV 683
 DB 833 RNKRGVYTRLPFLFDWIKENTGV 855
 RESULT 7
 ID AD116876
 XX AD116876 standard; protein: 855 AA.
 XX
 AC AD116876;
 XX
 DT 15-APR-2004 (first entry)
 XX

Db 533 SNGKCLSKSQCGKNGDKGSDGDEASCPKXNVVYCTKHTYRCLNGLCLSKGNPECDKED 592
Qy 421 CSDSDSDKDCGCGARSTFRQARVVGCTADGEMPMQVSLHALGQGHICGASLISPMNLV 480
Db 593 CSDSDSDKDCGCGARSTFRQARVVGCTADGEMPMQVSLHALGQGHICGASLISPMNLV 652
Qy 481 SAACGYIDDRGFRYSPTQMTAFGLHDSQSRAPGVGERLKRILISHPFNDFTFDYDI 540
Db 653 SAACGYIDDRGFRYSPTQMTAFGLHDSQSRAPGVGERLKRILISHPFNDFTFDYDI 712
Qy 541 ALLELEKPAEYSSWVRPICLPDASHVFPAGKAIWVGWGHITQYGGTGAIILOKGEIRVIN 600
Db 713 ALLELEKPAEYSSWVRPICLPDASHVFPAGKAIWVGWGHITQYGGTGAIILOKGEIRVIN 772
Qy 601 OTTGENLLPQGITRRMVCVGLSGVDSGCGDSSGRLSSVADGRIFQAGVWSMGDCAQ 660
Db 773 OTTGENLLPQGITRRMVCVGLSGVDSGCGDSSGRLSSVADGRIFQAGVWSMGDCAQ 832
Qy 661 RNKGVYTRLPFRDMIKENTGV 683
Db 833 RNKGVYTRLPFRDMIKENTGV 855

RESULT 8
ADN39867
ID ADN39867 standard; protein; 855 AA.
XX
AC ADN39867;
XX
DT 17-JUN-2004 (first entry)
XX
DE Cancer/angiogenesis/fibrosis-related polypeptide, SEQ ID NO: C237.
XX
XX Human, differential expression; cancer; angiogenic disorder;
KM fibrotic disorder; psoriasis; ischaemia; heart disease; atherosclerosis;
KM inflammatory disease; autoimmune disease;
KM retinal neovascularization syndrome; scarring; uterine fibroid;
KM detection; diagnosis; prognosis; drug screening; drug targeting;
KM wound healing; contraception; cytostatic; cardiac; immunomodulatory;
KM vulnery; gene therapy; vaccine.
XX
OS Homo sapiens.
XX
PN W02003042661-A2.
XX
PD 22-MAY-2003.
XX
PF 13-NOV-2002; 2002WO-US036810.
XX
PR 13-NOV-2001; 2001US-0350666P.
PR 21-NOV-2001; 2001US-0332464P.
PR 29-NOV-2001; 2001US-0334393P.
PR 03-DEC-2001; 2001US-0335394P.
PR 14-DEC-2001; 2001US-0340376P.
PR 08-JAN-2002; 2002US-0347211P.
PR 10-JAN-2002; 2002US-0347349P.
PR 08-FEB-2002; 2002US-0355250P.
PR 13-FEB-2002; 2002US-0356714P.
PR 20-FEB-2002; 2002US-0359077P.
PR 29-MAR-2002; 2002US-0368099P.
PR 04-APR-2002; 2002US-0370110P.
PR 12-APR-2002; 2002US-0372246P.
PR 05-JUN-2002; 2002US-0386614P.
PR 16-JUN-2002; 2002US-0396839P.
PR 22-JUL-2002; 2002US-0397755P.
PR 22-JUL-2002; 2002US-0397845P.
PR 09-SEP-2002; 2002US-0409450P.
XX
PA (EOSB-) EOS BIOTECHNOLOGY INC.
XX
XX Afar D, Aziz N, Ginsburg WM, Gish KC, Glynn R, Hevezi PA;
PI Mack DH, Murray R, Watson SR, Wilson KE, Zlotnik A;
XX

DR WPI: 2003-468649/44.
XX N-PSDB; ADN39650.
XX
PT Determining the presence or absence of a pathological cell in a patient,
PT useful for diagnosing, prognosing or treating cancer, comprises detecting
PT a nucleic acid in a biological sample.
XX
PS Claim 12; SEQ ID NO C237; 1385bp; English.
XX
XX The invention relates to nucleic acids and proteins (ADN38683-ADN40064)
CC whose expression is upregulated or downregulated in specific cancers or
CC other diseases such as angiogenic or fibrotic disorders, and to methods
CC of determining the presence or absence of a pathological cell in a
CC patient by detecting a nucleic acid at least 80% identical to those of
CC the invention or by detecting a polypeptide of the invention. The
CC invention also relates to expression vectors and host cells comprising a
CC nucleic acid of the invention; antibodies which specifically bind a
CC polypeptide of the invention; use of such antibodies for drug targeting;
CC and methods of screening for modulators of activity or expression of the
CC polypeptides and nucleic acids. The nucleic acids, polypeptides,
CC antibodies and methods are useful for diagnosing, prognosing and treating
CC cancer and other conditions such as psoriasis, ischaemia, heart disease,
CC atherosclerosis, inflammatory diseases, autoimmune diseases, retinal
CC neovascularization syndromes, scarring and uterine fibroids. They may
CC also be useful in wound healing and in contraception. The present
CC sequence represents a polypeptide of the invention.
XX
SQ Sequence 855 AA;
Query Match 99.2%; Score 3760; DB 7; Length 855;
Best Local Similarity 99.4%; Pred. No. 4.7e-241;
Matches 679; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 1 MAERVVMLPPRARSLKSFVYTSVVAPEPTSKYQRTQDNSCSFGHARVEMRFTTPG 60
Db 173 MAERVVMLPPRARSLKSFVYTSVVAPEPTSKYQRTQDNSCSFGHARVEMRFTTPG 232
Qy 61 PPDSPYPAHARCOMALGDDSVSLTFRRSFDLASGDERGSDLVYVNTLSPMEPHALVQ 120
Db 233 PPDSPYPAHARCOMALGDDSVSLTFRRSFDLASGDERGSDLVYVNTLSPMEPHALVQ 292
Qy 121 LCGTYPPSYNLTFHSSQNVLLITLITNTERRRHGFEATFPOLPMSCGRLRKAQGTFN 180
Db 293 LCGTYPPSYNLTFHSSQNVLLITLITNTERRRHGFEATFPOLPMSCGRLRKAQGTFN 352
Qy 181 SPYYPGHYPPNIDCTWNIIEVPNNQHVVRKFFYLEPRAAGTCCPDYVYINEKXCGE 240
Db 353 SPYYPGHYPPNIDCTWNIIEVPNNQHVVRKFFYLEPRAAGTCCPDYVYINEKXCGE 412
Qy 241 RSQFVYVTSNKNKTIVRHSQSYTDPDGLAEVLSYDSDSCPDQFCTORTGCTRKELRCD 300
Db 413 RSQFVYVTSNKNKTIVRHSQSYTDPDGLAEVLSYDSDSCPDQFCTORTGCTRKELRCD 472
Qy 301 GMAIDCTHSDDELNCSGAGHQFTCKNFKCKPLFWVCDVANDCGNSDEGCSGPAQTFRG 360
Db 473 GMAIDCTHSDDELNCSGAGHQFTCKNFKCKPLFWVCDVANDCGNSDEGCSGPAQTFRG 532
Qy 361 SNGKCLSKSQCGKNGDKGSDGDEASCPKXNVVYCTKHTYRCLNGLCLSKGNPECDKED 420
Db 533 SNGKCLSKSQCGKNGDKGSDGDEASCPKXNVVYCTKHTYRCLNGLCLSKGNPECDKED 592
Qy 421 CSDSDSDKDCGCGARSTFRQARVVGCTADGEMPMQVSLHALGQGHICGASLISPMNLV 480
Db 593 CSDSDSDKDCGCGARSTFRQARVVGCTADGEMPMQVSLHALGQGHICGASLISPMNLV 652
Qy 481 SAACGYIDDRGFRYSPTQMTAFGLHDSQSRAPGVGERLKRILISHPFNDFTFDYDI 540
Db 653 SAACGYIDDRGFRYSPTQMTAFGLHDSQSRAPGVGERLKRILISHPFNDFTFDYDI 712
Qy 541 ALLELEKPAEYSSWVRPICLPDASHVFPAGKAIWVGWGHITQYGGTGAIILOKGEIRVIN 600
Db 713 ALLELEKPAEYSSWVRPICLPDASHVFPAGKAIWVGWGHITQYGGTGAIILOKGEIRVIN 772

QY 601 QTTCEMLLPQOITPRMNCVGLSGVDSCQDSGGPLSVYADGRIFQAGVSWGDCAQ 660
DB 773 QTTCEMLLPQOITPRMNCVGLSGVDSCQDSGGPLSVYADGRIFQAGVSWGDCAQ 832
QY 661 RNKRGVYTRLPFRDWIKENTGV 683
DB 833 RNKRGVYTRLPFRDWIKENTGV 855

RESULT 9

ID ADN04754 standard; protein; 855 AA.
ADN04754;
AC ADN04754;
XX
XX
XX

DT 01-JUN-2004 (first entry)
DE Antipsoriatic protein sequence #558.
XX

XX antipsoriatic; gene therapy; psoriasis; diagnosis.
KM
XX
OS Homo sapiens.
XX

PN W02004028479-A2.
XX
XX
XX

PD 08-APR-2004.
XX
XX
XX

PF 25-SEP-2003; 2003WO-US030907.
XX
XX
XX

PR 25-SEP-2002; 2002US-0414006P.
XX
XX
XX

PA (GERTH) GENENTECH INC.
XX
XX
XX

PI Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WJ,
Wu TD;
XX
XX
XX

DR WPI; 2004-305105/28.
XX
XX
XX

DR N-PSDB; ADN04753.
XX
XX
XX

PT New PRO nucleic acid or polypeptide, useful for preparing a
pharmaceutical composition for diagnosing or treating psoriasis in a
mammal.
XX
XX
XX

PS Claim 9; SEQ ID NO 1148; 3063pp; English.
XX
XX
XX

CC The invention relates to novel polynucleotide and polypeptides for
treating psoriasis or a sequence having at least 80% identity to the
above sequences. The nucleic acid is useful for preparing a composition
for diagnosing or treating psoriasis in a mammal. This sequence
corresponds to one of the polypeptides of the invention.
XX
XX
XX

XX Sequence 855 AA;
XX
XX
XX

Query Match 99.2%; Score 3760; DB 8; Length 855;
Best Local Similarity 99.4%; Pred. No. 4,7e-241;
Matches 679; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 MAERVVMLPPRRASLKFVTVSVARPTDSKTVOQTODNSCSFGLARGVLMRFTTPG 60
DB 173 MAERVVMLPPRRASLKFVTVSVARPTDSKTVOQTODNSCSFGLARGVLMRFTTPG 232

QY 61 FPDSPYAHARCOMALRGDADSVSLTFRSFDLASCDERGSGLTVYNTLSPMEPHALVQ 120
DB 233 FPDSPYAHARCOMALRGDADSVSLTFRSFDLASCDERGSGLTVYNTLSPMEPHALVQ 292

QY 121 LCGTTPPSYNTLTFHSSQNVLLITLTTERRHGFEXATFFQLPRMSSCGGLRKAQGTN 180
DB 293 LCGTTPPSYNTLTFHSSQNVLLITLTTERRHGFEXATFFQLPRMSSCGGLRKAQGTN 352

QY 181 SPVYRGHYPPNIDCTWNIIEVNNQHVVRPFKFFLLPRRACGCPDYYEINEXYCGE 240
DB 353 SPVYRGHYPPNIDCTWNIIEVNNQHVVRPFKFFLLPRRACGCPDYYEINEXYCGE 412

QY 241 RSQFVTVSNMKTITVRFHSDQSYTDGFLAEVLSYDSDPCPOFTCRTGRICRKELRCD 300
DB 413 RSQFVTVSNMKTITVRFHSDQSYTDGFLAEVLSYDSDPCPOFTCRTGRICRKELRCD 472
QY 301 GMADCTDHSDELNCSCDAGHQFTCKNKFCKPLFWVCDVNDCCGDNBDEQCSGPAQTFRC 360
DB 473 GMADCTDHSDELNCSCDAGHQFTCKNKFCKPLFWVCDVNDCCGDNBDEQCSGPAQTFRC 532
QY 361 SNGKCLSKSQQCNKGDCCGDSDEASCPRNVVVTCTKTYRCNLGCLSKGNBECQKED 420
DB 533 SNGKCLSKSQQCNKGDCCGDSDEASCPRNVVVTCTKTYRCNLGCLSKGNBECQKED 592
QY 421 CSDGSEKDCDGLRSFTROARVVGTDADGEMPMQVSIHALGQGHICASLISPMWLV 480
DB 593 CSDGSEKDCDGLRSFTROARVVGTDADGEMPMQVSIHALGQGHICASLISPMWLV 652
QY 481 SAACHYTDGFRFYSPTQWTAFLGLHDSQRAPGVQERLRRIISHPEFNDFTDYDI 540
DB 653 SAACHYTDGFRFYSPTQWTAFLGLHDSQRAPGVQERLRRIISHPEFNDFTDYDI 712
QY 541 ALLELEKPAEYSMWBPICLPDASHVPAGKAIWVTGMGHTQYGGTALILQKGEIRVIN 600
DB 713 ALLELEKPAEYSMWBPICLPDASHVPAGKAIWVTGMGHTQYGGTALILQKGEIRVIN 772
QY 601 QTTCEMLLPQOITPRMNCVGLSGVDSCQDSGGPLSVYADGRIFQAGVSWGDCAQ 660
DB 773 QTTCEMLLPQOITPRMNCVGLSGVDSCQDSGGPLSVYADGRIFQAGVSWGDCAQ 832
QY 661 RNKRGVYTRLPFRDWIKENTGV 683
DB 833 RNKRGVYTRLPFRDWIKENTGV 855

RESULT 10

ID ADI16884 standard; protein; 855 AA.
ADI16884;
XX
XX
XX

DT 15-APR-2004 (first entry)
DE Human NOXV protein homologue Segid 420.
XX
XX
XX

KM human; NOXV; cardiomyopathy; atherosclerosis; cancer; diabetes;
inflammation; autoimmune disorder; allergy; blood disorder;
acquired immunodeficiency syndrome; AIDS; obesity; asthma;
immunoglobulin (Ig)A nephropathy; cirrhosis; arthritis;
Alzheimer's disease; infection; str.
XX
XX
XX

OS Homo sapiens.
XX
XX
XX

PN W0200268649-A2.
XX
XX
XX

PD 06-SEP-2002.
XX
XX
XX

PF 31-JAN-2002; 2002WO-US002785.
XX
XX
XX

PR 31-JAN-2001; 2001US-0265395P.
XX
XX
XX

PR 31-JAN-2001; 2001US-0265412P.
XX
XX
XX

PR 31-JAN-2001; 2001US-0265514P.
XX
XX
XX

PR 02-FEB-2001; 2001US-0266406P.
XX
XX
XX

PR 05-FEB-2001; 2001US-0266767P.
XX
XX
XX

PR 07-FEB-2001; 2001US-0266757P.
XX
XX
XX

PR 08-FEB-2001; 2001US-0267057P.
XX
XX
XX

PR 09-FEB-2001; 2001US-0267459P.
XX
XX
XX

PR 15-FEB-2001; 2001US-0267823P.
XX
XX
XX

PR 26-FEB-2001; 2001US-0271664P.
XX
XX
XX

PR 27-FEB-2001; 2001US-0271839P.
XX
XX
XX

PR 02-MAR-2001; 2001US-0272788P.
XX
XX
XX

PR 14-MAR-2001; 2001US-0275925P.
PR 14-MAR-2001; 2001US-0275947P.
PR 14-MAR-2001; 2001US-0275950P.
PR 14-MAR-2001; 2001US-0275989P.
PR 15-MAR-2001; 2001US-0276448P.
PR 15-MAR-2001; 2001US-0276450P.
PR 16-MAR-2001; 2001US-0276397P.
PR 16-MAR-2001; 2001US-0276768P.
PR 20-MAR-2001; 2001US-0276852P.
PR 26-MAR-2001; 2001US-0278775P.
PR 26-MAR-2001; 2001US-0278778P.
PR 29-MAR-2001; 2001US-0279882P.
PR 29-MAR-2001; 2001US-0279884P.
PR 30-MAR-2001; 2001US-0280147P.
PR 11-APR-2001; 2001US-0282992P.
PR 11-APR-2001; 2001US-0283083P.
PR 20-APR-2001; 2001US-0285133P.
PR 23-APR-2001; 2001US-0285749P.
PR 03-MAY-2001; 2001US-0288327P.
PR 03-MAY-2001; 2001US-0288504P.
PR 29-MAY-2001; 2001US-0294047P.
PR 30-MAY-2001; 2001US-0294473P.
PR 08-JUN-2001; 2001US-0296964P.
PR 18-JUN-2001; 2001US-0298959P.
PR 19-JUN-2001; 2001US-0312020P.
PR 16-AUG-2001; 2001US-0312889P.
PR 16-AUG-2001; 2001US-0312908P.
PR 21-AUG-2001; 2001US-0313390P.
PR 28-AUG-2001; 2001US-0315470P.
PR 31-AUG-2001; 2001US-0316447P.
PR 07-SEP-2001; 2001US-0318115P.
PR 07-SEP-2001; 2001US-0318118P.
PR 12-SEP-2001; 2001US-0318740P.
PR 19-SEP-2001; 2001US-0323379P.
PR 18-OCT-2001; 2001US-0330245P.
PR 18-OCT-2001; 2001US-0330308P.
PR 14-NOV-2001; 2001US-0332701P.
XX
XX (CURA-) CURAGEN CORP.
XX Tcherven VT, Spytek KA, Zernhusen BD, Paturajan M, Shinkets RA;
PI Li L, Gargolli EA, Padigaru M, Anderson DM, Rastelli L, Miller CB;
PI Gelach V, Taudier RJ, Gusev VY, Colman SD, Wolenci AR, Pena CE;
PI Furtak K, Groese WM, Alsebrook JP, Lepley DM, Rieger DK, Burgess CE;
XX WPI; 2002-706998/76.
XX
XX New NOXV polypeptides and nucleic acids, useful for preventing or
PT treating NOXV-associated disorders, e.g. cancer, cardiomyopathy,
PT atherosclerosis, or diabetes, and in chromosome mapping, tissue typing or
PT pharmacogenomics.
XX
XX Disclosure; SEQ ID NO 420; 1498bp; English.
XX
XX This invention relates to a novel nucleic acid, and encoded polypeptides
CC thereof, which have properties related to the stimulation of biochemical
CC or physiological responses in a cell, tissue, organ or organism.
CC Specifically, it refers to the use of biologically active fragments for
CC diagnostic and prognostic assays and furthermore in the treatment of
CC diverse pathological conditions. The present invention describes novel
CC human and murine NOXV proteins, as well as methods to modulate their
CC expression using antisense oligos, ribozymes and peptide nucleic acids.
CC The NOXV polypeptides, polynucleotides and antibodies are useful in
CC treating or preventing NOXV-associated disorders, e.g. cardiomyopathy,
CC atherosclerosis, cancer and diabetes. Furthermore, they may be used in
CC treating or preventing diseases such as inflammation, autoimmune
CC disorders, allergies, blood disorders, acquired immunodeficiency syndrome
CC (AIDS), obesity, asthma, immunoglobulin (Ig)A nephropathy, cirrhosis,
CC arthritis, Alzheimer's disease, infections, stroke, muscular dystrophy
CC and epilepsy. Accordingly, these molecules have many activities including
CC cytoskeletal, cardiac, antiinflammatory, immunosuppressive, antiallergic,
CC haemostatic, anti-HIV, antidiabetic, antiarteriosclerotic, anorectic,

CC antiaesthetic, nephrotropic, antiarthritic, hepatotropic,
CC neuroprotective, nootropic, antibacterial, vitricide, antiparasitic,
CC relaxant and anticonvulsant. In addition, they are useful in screening
CC assays to identify small molecules that modulate or inhibit, for example,
CC neurogenesis, wound healing and angiogenesis. The nucleic acids are also
CC used as in chromosome mapping, tissue typing, preventive medicine and
CC pharmacogenomics. This polypeptide is a homologue of a human NOXV protein
CC of the invention.
XX
XX Sequence 855 AA:
SQ
Query Match 99.1%; Score 3756; DB 5; Length 855;
Best Local Similarity 99.3%; Pred. No. 8,6e-241;
Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 MAERVVMLEPPRASLSFVVTSSVAAPPTDSKTYQRTQDNCSTGLHARGVELMRFTTPG 60
DB 173 MAERVVMLEPPRASLSFVVTSSVAAPPTDSKTYQRTQDNCSTGLHARGVELMRFTTPG 232
QY 61 PPDSPPYPAHARCOMALRGDADSVSLTFRSPDLASCDERGSDLVTYVNTLSPMEPALVQ 120
DB 233 PPDSPPYPAHARCOMALRGDADSVSLTFRSPDLASCDERGSDLVTYVNTLSPMEPALVQ 292
QY 121 LCGTTPPSYNLTFSSQNVLLITLITNTERRHPEFATFPQLPRMSSCGRLRAQGTFN 180
DB 293 LCGTTPPSYNLTFSSQNVLLITLITNTERRHPEFATFPQLPRMSSCGRLRAQGTFN 352
QY 181 SPYYPGHYPNIDCTWNIETVPPNNOHVYRFEFFLLBPRACGTCPPDYVINEKTCGE 240
DB 353 SPYYPGHYPNIDCTWNIETVPPNNOHVYRFEFFLLBPRACGTCPPDYVINEKTCGE 412
QY 241 RSQFVMTNSNKKIVRFPSSDSTYDTGFLAEVLSYDSDPCPGQFTCRGTCIRKELRCD 300
DB 413 RSQFVMTNSNKKIVRFPSSDSTYDTGFLAEVLSYDSDPCPGQFTCRGTCIRKELRCD 472
QY 301 GMAADCTHSDLANCSCDAGHFTCKNFKCPLFWVCDSVNDGSDNSDOGSCSPAQTFCR 360
DB 473 GMAADCTHSDLANCSCDAGHFTCKNFKCPLFWVCDSVNDGSDNSDOGSCSPAQTFCR 532
QY 361 SNGKCLSKSQCKGCKGDCGSDASCPKXNVVTCTHTYRCLNGLCLSKNPECDGED 420
DB 533 SNGKCLSKSQCKGCKGDCGSDASCPKXNVVTCTHTYRCLNGLCLSKNPECDGED 592
QY 421 CSDSDSDKDCGGRSTRTQARVYVGTADGEMPOVSLHAGQGHICGASLISPMVLV 480
DB 593 CSDSDSDKDCGGRSTRTQARVYVGTADGEMPOVSLHAGQGHICGASLISPMVLV 652
QY 481 SAACHYIDDRGFRYSDEPTOWTAFGLHDQSRASAPGQERRLRKRIISHPFENDETFDYDI 540
DB 653 SAACHYIDDRGFRYSDEPTOWTAFGLHDQSRASAPGQERRLRKRIISHPFENDETFDYDI 712
QY 541 ALLELEKPAEYSSWVRPCLCPDASHVPAGKAIWVTGWHGTQYCGTALLIOKEIRVIN 600
DB 713 ALLELEKPAEYSSWVRPCLCPDASHVPAGKAIWVTGWHGTQYCGTALLIOKEIRVIN 772
QY 601 OTTCENLILPOITIRMMCVGFLSGVDSCGDSGSPSYEADRIROAGVSGDCAQ 660
DB 773 OTTCENLILPOITIRMMCVGFLSGVDSCGDSGSPSYEADRIROAGVSGDCAQ 832
QY 661 RNKPGVYTRLPFLFDWKENTGV 683
DB 833 RNKPGVYTRLPFLFDWKENTGV 855
RESULT 11
AD116818
ID AD116818 standard; protein; 855 AA.
XX
XX AD116818;
XX
XX 15-APR-2004 (first entry)
XX
XX Human NOXV protein homologue Segid 354.

XX human; NOX; cardiomyopathy; atherosclerosis; cancer; diabetes;
KW inflammation; autoimmune disorder; allergy; blood disorder;
KW acquired immunodeficiency syndrome; AIDS; obesity; asthma;
KW immunoglobulin (Ig)A nephropathy; cirrhosis; arthritis;
KW Alzheimer's disease; infection; str.
XX Homo sapiens.
XX
XX WO200268649-A2.
XX
XX 06-SEP-2002.
XX
XX 31-JAN-2002; 2002WO-US002785.
XX
XX 31-JAN-2001; 2001US-0265395P.
XX 31-JAN-2001; 2001US-0265412P.
XX 31-JAN-2001; 2001US-0265514P.
XX 31-JAN-2001; 2001US-0265517P.
XX 02-FEB-2001; 2001US-0266406P.
XX 05-FEB-2001; 2001US-0266767P.
XX 07-FEB-2001; 2001US-0266975P.
XX 07-FEB-2001; 2001US-0267057P.
XX 08-FEB-2001; 2001US-0267459P.
XX 09-FEB-2001; 2001US-0267823P.
XX 15-FEB-2001; 2001US-0268974P.
XX 26-FEB-2001; 2001US-0271664P.
XX 27-FEB-2001; 2001US-0271839P.
XX 27-FEB-2001; 2001US-0271855P.
XX 02-MAR-2001; 2001US-0272788P.
XX 02-MAR-2001; 2001US-0273046P.
XX 14-MAR-2001; 2001US-0275925P.
XX 14-MAR-2001; 2001US-0275947P.
XX 14-MAR-2001; 2001US-0275950P.
XX 14-MAR-2001; 2001US-0275989P.
XX 15-MAR-2001; 2001US-0276488P.
XX 15-MAR-2001; 2001US-0276450P.
XX 16-MAR-2001; 2001US-0276379P.
XX 16-MAR-2001; 2001US-0276768P.
XX 20-MAR-2001; 2001US-0278652P.
XX 26-MAR-2001; 2001US-0278775P.
XX 26-MAR-2001; 2001US-0278778P.
XX 29-MAR-2001; 2001US-0279882P.
XX 29-MAR-2001; 2001US-0279884P.
XX 30-MAR-2001; 2001US-0280147P.
XX 11-APR-2001; 2001US-0282992P.
XX 11-APR-2001; 2001US-0283083P.
XX 20-APR-2001; 2001US-0285133P.
XX 23-APR-2001; 2001US-0285749P.
XX 03-MAY-2001; 2001US-0288327P.
XX 03-MAY-2001; 2001US-0288504P.
XX 29-MAY-2001; 2001US-0294047P.
XX 30-MAY-2001; 2001US-0294473P.
XX 08-JUN-2001; 2001US-0296964P.
XX 18-JUN-2001; 2001US-0298959P.
XX 19-JUN-2001; 2001US-0299324P.
XX 13-AUG-2001; 2001US-0312020P.
XX 16-AUG-2001; 2001US-0312889P.
XX 16-AUG-2001; 2001US-0312908P.
XX 21-AUG-2001; 2001US-0313380P.
XX 28-AUG-2001; 2001US-0315470P.
XX 31-AUG-2001; 2001US-0316447P.
XX 07-SEP-2001; 2001US-0318115P.
XX 07-SEP-2001; 2001US-0318118P.
XX 12-SEP-2001; 2001US-0318740P.
XX 19-SEP-2001; 2001US-0323379P.
XX 18-OCT-2001; 2001US-0330245P.
XX 18-OCT-2001; 2001US-0330308P.
XX 14-NOV-2001; 2001US-0332701P.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Tchernev VT, Spytek KA, Zetthusen BD, Patturajan M, Shimkets RA;
PI

PI Li L, Gangoli EA, Padigaru M, Anderson DW, Rastelli L, Miller CE;
PI Gerlach VL, Taupier RJ, Gusev VY, Colman SD, Wolenc AR, Pena CBA;
PI Futrak K, Grosse WM, Alsobrook JP, Lopley DM, Rieger DK, Burgess CF;
XX
XX WPI; 2002-706998/76.
XX
XX
XX New NOX polypeptides and nucleic acids, useful for preventing or
PT treating NOX-associated disorders, e.g. cancer, cardiomyopathy,
PT atherosclerosis, or diabetes, and in chromosome mapping, tissue typing or
PT pharmacogenomics.
XX
XX
XX Disclosure; SEQ ID NO 354; 1498pp; English.
XX
XX
XX This invention relates to a novel nucleic acids, and encoded polypeptides
CC thereof, which have properties related to the stimulation of biochemical
CC or physiological responses in a cell, tissue, organ or organism.
CC Specifically, it refers to the use of biologically active fragments for
CC diagnostic and prognostic assays and furthermore in the treatment of
CC diverse pathological conditions. The present invention describes novel
CC human and murine NOX proteins, as well as methods to modulate their
CC expression using antisense oligos, ribozymes and peptide nucleic acids.
CC The NOX polypeptides, polynucleotides and antibodies are useful in
CC treating or preventing NOX-associated disorders, e.g. cardiomyopathy,
CC atherosclerosis, cancer and diabetes. Furthermore, they may be used in
CC treating or preventing diseases such as inflammation, autoimmune
CC disorders, allergies, blood disorders, acquired immunodeficiency syndrome
CC (AIDS), obesity, asthma, immunoglobulin (Ig)A nephropathy, cirrhosis,
CC arthritis, Alzheimer's disease, infections, stroke, muscular dystrophy
CC and epilepsy. Accordingly, these molecules have many activities including
CC cytostatic, cardiant, anti-inflammatory, immunosuppressive, antiallergic,
CC haemostatic, anti-HIV, antidiabetic, antiarteriosclerotic, anorectic,
CC antistatic, nephrotropic, antiarthritic, hepatotropic,
CC neuroprotective, neurotropic, antibacterial, virucide, antiparasitic,
CC relaxant and anticoagulant. In addition, they are useful in screening
CC assays to identify small molecules that modulate or inhibit, for example,
CC neurogenesis, wound healing and angiogenesis. The nucleic acids are also
CC used as in chromosome mapping, tissue typing, preventive medicine and
CC pharmacogenomics. This polypeptide is a homologue of a human NOX protein
XX of the invention.
XX
XX
XX Sequence 855 AA;
SQ
Query Match 99.1%; Score 3756; DB 5; Length 855;
Best Local Similarity 99.3%; Pred. No. 8, 6e-241;
Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 MAEERVVMLPPRARSLKSPFVTSVVAFPDTSKVYORTQONSCEFGHARGVEMRFTTPG 60
DB 173 MAEERVVMLPPRARSLKSPFVTSVVAFPDTSKVYORTQONSCEFGHARGVEMRFTTPG 232
QY 61 FPDSPYPAHARCOMALRGADSVLSTLFRSFDLASCDERSGDLVTVYNTLSPEMHALVQ 120
DB 233 FPDSPYPAHARCOMALRGADSVLSTLFRSFDLASCDERSGDLVTVYNTLSPEMHALVQ 232
QY 121 LCGTTPSYNLTFHSSQNVLLITLINTERRHPGFATFFQLPRMSSCGRLKAGCTFN 180
DB 293 LCGTTPSYNLTFHSSQNVLLITLINTERRHPGFATFFQLPRMSSCGRLKAGCTFN 352
QY 181 SPYYPGHYPNDICTNNINFPNNOHVAFKFPYLLPBRACGTCCKDYVEINGEYCCG 240
DB 353 SPYYPGHYPNDICTNNINFPNNOHVAFKFPYLLPBRACGTCCKDYVEINGEYCCG 412
QY 241 RSQFVVTNSNKTITVAFHSDOSYTDGFLAELYLSYDSDPCPGQFTCRGRCIRKELRCD 300
DB 413 RSQFVVTNSNKTITVAFHSDOSYTDGFLAELYLSYDSDPCPGQFTCRGRCIRKELRCD 472
QY 301 GWADCTDHSDELNCSCDAGHOFCKKFKCPLFWVCDSDVNDGDNBDEGCSCPAOTFRG 360
DB 473 GWADCTDHSDELNCSCDAGHOFCKKFKCPLFWVCDSDVNDGDNBDEGCSCPAOTFRG 532
QY 361 SNGKCLSKSQGCKXDCGSDGSDASCPKYNVYTCRKHYYRCLNGLCKSGNPEGCKGKED 420
DB 533 SNGKCLSKSQGCKXDCGSDGSDASCPKYNVYTCRKHYYRCLNGLCKSGNPEGCKGKED 592

QY 421 CSDGSEKDCDCGLRSFTROARVVGSTPDADGEMPMOYSLHALGQGHICGASLISPMNLV 480
 DB 593 CSDGSEKDCDCGLRSFTROARVVGSTPDADGEMPMOYSLHALGQGHICGASLISPMNLV 652
 QY 481 SAACHYIDRGRFRYSPTQMTAFGLHDSQSRAPGVORRLKRISHPFNDFTFYDI 540
 DB 653 SAACHYIDRGRFRYSPTQMTAFGLHDSQSRAPGVORRLKRISHPFNDFTFYDI 712
 QY 541 ALLELEKAEYSSMWRIPLDPASHVFPAGKAIWVTGNGHGYGGTGALLIQKEIRYIN 600
 DB 713 ALLELEKAEYSSMWRIPLDPASHVFPAGKAIWVTGNGHGYGGTGALLIQKEIRYIN 772
 QY 601 OTTCENILPQOITPPMMCVGFLSGVDSQQDSSGGLSVVADGRIPGAGVSMGDCGQ 660
 DB 773 OTTCENILPQOITPPMMCVGFLSGVDSQQDSSGGLSVVADGRIPGAGVSMGDCGQ 832
 QY 661 RNKPGVYTRLPFRDWIKENTGV 683
 DB 833 RNKPGVYTRLPFRDWIKENTGV 855

RESULT 12
 AD116508
 ID AD116508 standard; protein: 757 AA.
 XX
 AC AD116508;
 DT 15-APR-2004 (first entry)
 XX
 DE Human NOVX protein to treat human pathological conditions SeqId4.
 XX
 KW human; NOVX; cardiomyopathy; atherosclerosis; cancer; diabetes;
 KW inflammation; autoimmune disorder; allergy; blood disorder;
 KW acquired immunodeficiency syndrome; AIDS; obesity; asthma;
 KW immunoglobulin (Ig)A nephropathy; cirrhosis; arthritis;
 KW Alzheimer's disease; infection; stroke; muscular dystrophy; epilepsy;
 KW cytosarctic; cardiac; antiinflammatory; immunosuppressive; antiallergic;
 KW haemostatic; anti-HIV; antidiabetic; antiarteriosclerotic; anorectic;
 KW antiaethmatic; nephrotropic; antiarthritic; hepatotropic;
 KW neuroprotective; nootropic; antibacterial; virucide; antiparasitic;
 KW relaxant; anticonvulsant; neurogenesis; wound healing; angiogenesis;
 KW chromosome mapping; tissue typing; pharmacogenomic.
 XX
 OS Homo sapiens.
 XX
 XX WO200268649-A2.
 XX
 PD 06-SEP-2002.
 XX
 PF 31-JAN-2002; 2002WO-US002785.
 XX
 PR 31-JAN-2001; 2001US-0265395P.
 PR 31-JAN-2001; 2001US-0265412P.
 PR 31-JAN-2001; 2001US-0265514P.
 PR 31-JAN-2001; 2001US-0265517P.
 PR 02-FEB-2001; 2001US-0266406P.
 PR 05-FEB-2001; 2001US-0266767P.
 PR 07-FEB-2001; 2001US-0266975P.
 PR 08-FEB-2001; 2001US-0267057P.
 PR 09-FEB-2001; 2001US-0267823P.
 PR 15-FEB-2001; 2001US-0268974P.
 PR 26-FEB-2001; 2001US-0271664P.
 PR 27-FEB-2001; 2001US-0271833P.
 PR 27-FEB-2001; 2001US-0271855P.
 PR 02-MAR-2001; 2001US-0272788P.
 PR 02-MAR-2001; 2001US-0273046P.
 PR 14-MAR-2001; 2001US-0275925P.
 PR 14-MAR-2001; 2001US-0275947P.
 PR 14-MAR-2001; 2001US-0275950P.
 PR 14-MAR-2001; 2001US-0275989P.
 PR 15-MAR-2001; 2001US-0276448P.

PR 15-MAR-2001; 2001US-0276450P.
 PR 16-MAR-2001; 2001US-0276397P.
 PR 16-MAR-2001; 2001US-0276768P.
 PR 20-MAR-2001; 2001US-0278652P.
 PR 26-MAR-2001; 2001US-0278775P.
 PR 26-MAR-2001; 2001US-0278778P.
 PR 29-MAR-2001; 2001US-0279882P.
 PR 29-MAR-2001; 2001US-0279884P.
 PR 30-MAR-2001; 2001US-0280147P.
 PR 11-APR-2001; 2001US-0282992P.
 PR 11-APR-2001; 2001US-0283083P.
 PR 20-APR-2001; 2001US-0285133P.
 PR 23-APR-2001; 2001US-0285749P.
 PR 03-MAY-2001; 2001US-0288327P.
 PR 03-MAY-2001; 2001US-0288504P.
 PR 29-MAY-2001; 2001US-0294047P.
 PR 30-MAY-2001; 2001US-0294473P.
 PR 08-JUN-2001; 2001US-0296964P.
 PR 18-JUN-2001; 2001US-0298959P.
 PR 19-JUN-2001; 2001US-0299324P.
 PR 13-AUG-2001; 2001US-0312020P.
 PR 16-AUG-2001; 2001US-0312889P.
 PR 16-AUG-2001; 2001US-0312908P.
 PR 21-AUG-2001; 2001US-0313390P.
 PR 28-AUG-2001; 2001US-0315470P.
 PR 31-AUG-2001; 2001US-0316447P.
 PR 07-SEP-2001; 2001US-0318115P.
 PR 07-SEP-2001; 2001US-0318118P.
 PR 12-SEP-2001; 2001US-0318740P.
 PR 19-SEP-2001; 2001US-0323379P.
 PR 18-OCT-2001; 2001US-0330245P.
 PR 18-OCT-2001; 2001US-0330308P.
 PR 14-NOV-2001; 2001US-0332701P.

(CURA-) CURAGEN CORP.
 XX
 PA Tcherenev VT, Spytek KA, Zernusen BD, Patturajan M, Shinkets RA;
 XX Li L, Gangolli EA, Padigan M, Anderson DM, Rastelli L, Miller CE;
 PI Gerlach V, Taupier RJ, Gusev VY, Colman SD, Wolenc AR, Pena CEA;
 PI Futak K, Grose WM, Alsbrook JP, Lepley DM, Rieger DK, Burgess CE;
 XX
 DR WPI; 2002-706998/76.
 XX N-PSDB; AD116507.
 XX
 PT New NOVX polypeptides and nucleic acids, useful for preventing or
 PT treating NOVX-associated disorders, e.g. cancer, cardiomyopathy,
 PT atherosclerosis, or diabetes, and in chromosome mapping, tissue typing or
 PT pharmacogenomics.
 XX
 PS Claim 1; SEQ ID NO 44; 1498bp; English.

This invention relates to a novel nucleic acids, and encoded polypeptides thereof, which have properties related to the stimulation of biochemical or physiological responses in a cell, tissue, organ or organism. Specifically, it refers to the use of biologically active fragments for diagnostic and prognostic assays and furthermore in the treatment of diverse pathological conditions. The present invention describes novel human and murine NOVX proteins, as well as methods to modulate their expression using antisense oligos, ribozymes and peptide nucleic acids. The NOVX polypeptides, polynucleotides and antibodies are useful in treating or preventing NOVX-associated disorders, e.g. cardiomyopathy, atherosclerosis, cancer and diabetes. Furthermore, they may be used in treating or preventing diseases such as inflammation, autoimmune disorders, allergies, blood disorders, acquired immunodeficiency syndrome (AIDS), obesity, asthma, immunoglobulin (Ig)A nephropathy, cirrhosis, arthritis, Alzheimer's disease, infections, stroke, muscular dystrophy and epilepsy. Accordingly, these molecules have many activities including cytosarctic, cardiac, antiinflammatory, immunosuppressive, antiallergic, haemostatic, anti-HIV, antidiabetic, antiarteriosclerotic, anorectic, antiaethmatic, nephrotropic, antiarthritic, hepatotropic, neuroprotective, nootropic, antibacterial, virucide, antiparasitic, relaxant and anticonvulsant. In addition, they are useful in screening assays to identify small molecules that modulate or inhibit, for example,

CC neurogenesis, wound healing and angiogenesis. The nucleic acids are also
 CC used as in chromosome mapping, tissue typing, preventive medicine and
 CC pharmacogenomics. This polypeptide is a human NOVX protein of the
 CC invention.

XX
 CC
 SQ Sequence 757 AA;

Query Match 99.1%; Score 3754; DB 5; Length 757;
 Best Local Similarity 99.3%; Pred. No. 1e-240;
 Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAERVVMLPPRASIKSFVVTSTVAAPPTDSKTVQRTQDNCSCFGLHARGVELMRFTTPG 60
 DB 75 MAERVVMLPPRASIKSFVVTSTVAAPPTDSKTVQRTQDNCSCFGLHARGVELMRFTTPG 134
 QY 61 FPDSPYPAHARCOMALRGDADSVSLTFRSFDLASCDERGSIDLVTYNTLSPMEPHALVQ 120
 DB 135 FPDSPYPAHARCOMALRGDADSVSLTFRSFDLASCDERGSIDLVTYNTLSPMEPHALVQ 194
 QY 121 LCGTPPSYNLTFHSSONVLLITLITNTERRHGFEATFPQLPRMSSCGRLRKAQGTEN 180
 DB 195 LCGTPPSYNLTFHSSONVLLITLITNTERRHGFEATFPQLPRMSSCGRLRKAQGTEN 254
 QY 181 SPYRPGHYPPNIDCTNMIENVNNOHYKRFKFFLLEPRACGTCPKDYVEINEKTCGE 240
 DB 255 SPYRPGHYPPNIDCTNMIENVNNOHYKRFKFFLLEPRACGTCPKDYVEINEKTCGE 314
 QY 241 RSQFVTVSNSSNKITVRFHSDSYTDGTFLAEYLSYSDSDPCPGQFTCRGTGCIKELRCD 300
 DB 315 RSQFVTVSNSSNKITVRFHSDSYTDGTFLAEYLSYSDSDPCPGQFTCRGTGCIKELRCD 374
 QY 301 GMADCTDHSDELNCSDAHQFTCKNKPCKPLFVWCDSDVNDGDSDEQGSCEPAQTFCR 360
 DB 375 GMADCTDHSDELNCSDAHQFTCKNKPCKPLFVWCDSDVNDGDSDEQGSCEPAQTFCR 434
 QY 361 SNGKCLSKSQOCCNGKDCGDSDEASCPKXNVVCTHRTYRCNLGLCLSKNPECDKED 420
 DB 435 SNGKCLSKSQOCCNGKDCGDSDEASCPKXNVVCTHRTYRCNLGLCLSKNPECDKED 494
 QY 421 CSDGSDERKDCDCGLRSFTQARVVGCTDADGEMPMQVSLHALQGHICGASLISPMNLV 480
 DB 495 CSDGSDERKDCDCGLRSFTQARVVGCTDADGEMPMQVSLHALQGHICGASLISPMNLV 554
 QY 481 SAACHYIDDRGFYSDFPTQWTAFLGLHDQSRASPAVGQERRLKRIISHPEFNDFTFYDI 540
 DB 555 SAACHYIDDRGFYSDFPTQWTAFLGLHDQSRASPAVGQERRLKRIISHPEFNDFTFYDI 614
 QY 541 ALLELEKPEAYSSMVRPICLPDASHVPAGKAIWVTGMHTQYCGTGALLIQKEIRVIN 600
 DB 615 ALLELEKPEAYSSMVRPICLPDASHVPAGKAIWVTGMHTQYCGTGALLIQKEIRVIN 674
 QY 601 QTTCENTLPPQITPRMMCVGFLSGVNSCOGDSGAPLSYEADGRIFQAGVSWGDCQAQ 660
 DB 675 QTTCENTLPPQITPRMMCVGFLSGVNSCOGDSGAPLSYEADGRIFQAGVSWGDCQAQ 734
 QY 661 RNKPGVYTRLLPLFRDWIKENTGV 683
 DB 735 RNKPGVYTRLLPLFRDWIKENTGV 757

RESULT 13
 AAY06671
 ID AAY06671 standard; protein; 855 AA.
 XX
 XX AAY06671;

XX 09-NOV-1999 (first entry)
 XX Tumour antigen derived gene-15 (TADG-15) protein.
 XX Tumour antigen derived gene-15; TADG-15; serine protease; human;
 KM breast cancer; ovary cancer; carcinoma; diagnosis.
 XX

OS Homo sapiens.

XX Key Location/Qualifiers
 FH Domain 1..54
 FT Domain /note= "cytoplasmic domain"
 FT Domain 55..213
 FT Modified-site /note= "transmembrane domain"
 FT Modified-site 109..111
 FT Region /note= "Aen is N-glycosylated"
 FT Region 214..447
 FT Region /note= "CUB repeat"
 FT Modified-site 302..304
 FT Region /note= "Aen is N-glycosylated"
 FT Region 453..602
 FT Region /note= "ligand-binding repeat (class A motif)"
 FT Region 481..483
 FT Region /note= "conserved SDE motif"
 FT Region 518..520
 FT Region /note= "conserved SDE motif"
 FT Region 554..556
 FT Region /note= "conserved SDE motif"
 FT Region 597..599
 FT Cleavage-site /note= "conserved SDE motif"
 FT Domain 614..615
 FT Domain 615..855
 FT /note= "catalytic domain"

XX WO9942120-A1.

XX 26-AUG-1999.

XX 18-FEB-1999; 99WO-US003436.

XX 20-FEB-1998; 98US-00027337.

XX (UYAR-) UNIV ARKANSAS.

XX O'brien TJ, Tanimoto H;

XX WPI; 1999-527418/44.

XX DR N-PSDB; AAX87815.

XX A new extracellular serine protease for diagnosis of neoplastic disease.

XX Claim 3; Fig 10; 71pp; English.

XX The present sequence represents a novel human extracellular serine
 CC protease, termed tumour antigen derived gene-15 protein (see AAY06671),
 CC that is overexpressed in breast and ovarian carcinomas. The TADG-15 gene
 CC (see AAX87815) can be used as a diagnostic and therapeutic target in
 CC ovarian carcinoma and other carcinomas including breast, prostate, lung
 CC and colon. The ligand binding domains of TADG-15 may be valuable in the
 CC uptake of specific molecules into tumour cells. The invention also
 CC provides: a vector that is capable of expressing DNA encoding TADG-15
 CC protein; host cells selected from bacterial cells (especially Escherichia
 CC coli), mammalian cells, plant cells and insect cells; and a method of
 CC detecting expression of TADG-15 protein using a hybridisation probe

XX Sequence 855 AA;

Query Match 99.1%; Score 3754; DB 2; Length 855;
 Best Local Similarity 99.3%; Pred. No. 1.2e-240;
 Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAERVVMLPPRASIKSFVVTSTVAAPPTDSKTVQRTQDNCSCFGLHARGVELMRFTTPG 60
 DB 173 MAERVVMLPPRASIKSFVVTSTVAAPPTDSKTVQRTQDNCSCFGLHARGVELMRFTTPG 232
 QY 61 FPDSPYPAHARCOMALRGDADSVSLTFRSFDLASCDERGSIDLVTYNTLSPMEPHALVQ 120
 DB 233 FPDSPYPAHARCOMALRGDADSVSLTFRSFDLASCDERGSIDLVTYNTLSPMEPHALVQ 292
 QY 121 LCGTPPSYNLTFHSSONVLLITLITNTERRHGFEATFPQLPRMSSCGRLRKAQGTEN 180

Db 293 LCGTPPSYNTLTFHSSQVLLITLITNTERHHPGEATFFQLPKMSSCGRLRKAQGTFN 352
Qy 181 SPYPGHYPNIDCTWNIIEVNNQHVKVRFPKPYLLBPRRACGTCPKDYVEINSEKTCGE 240
Db 353 SPYPGHYPNIDCTWNIIEVNNQHVKVSFPKPYLLBEGVAGTCPKDYVEINSEKTCGE 412
Qy 241 RSQFVTVSNKKTIVRFHSDSYDTGTGLAEYLSYDSDPCPGQFTCRGCRKELKCD 300
Db 413 RSQFVTVSNKKTIVRFHSDSYDTGTGLAEYLSYDSDPCPGQFTCRGCRKELKCD 472
Qy 301 GMACTDHSDELNCSDAHQFTCKNKECKPLFWVCDSDVNDGSDSGSCPAQTRFC 360
Db 473 GMACTDHSDELNCSDAHQFTCKNKECKPLFWVCDSDVNDGSDSGSCPAQTRFC 532
Qy 361 SNGKCLSKSQCGNKGDCGSDSDASCPKVVVVTCTKHTYRCLNGLCLSKNPECDGKED 420
Db 533 SNGKCLSKSQCGNKGDCGSDSDASCPKVVVVTCTKHTYRCLNGLCLSKNPECDGKED 592
Qy 421 CSDGSDXDCDCCGLRSFTQARVVGTDADGEMPMQVSLHALGQGHICGASLISPMWLV 480
Db 593 CSDGSDXDCDCCGLRSFTQARVVGTDADGEMPMQVSLHALGQGHICGASLISPMWLV 652
Qy 481 SAACIYIDDRGFRYSDFQWTAFLGLHDQSORSAFGVERLKRITISHPFNDFTFDYDI 540
Db 653 SAACIYIDDRGFRYSDFQWTAFLGLHDQSORSAFGVERLKRITISHPFNDFTFDYDI 712
Qy 541 ALLELEKPAEYSSMWRIICLPDASHVPAGKAIWVTGNGHTQYGGTALILQKEIRVIN 600
Db 713 ALLELEKPAEYSSMWRIICLPDASHVPAGKAIWVTGNGHTQYGGTALILQKEIRVIN 772
Qy 601 QTTCEMLLPQOITPRMNCVGLSGVDSGQSDSGPLSSVADGRIIFQAGVWSMGDCAQ 660
Db 773 QTTCEMLLPQOITPRMNCVGLSGVDSGQSDSGPLSSVADGRIIFQAGVWSMGDCAQ 832
Qy 661 RNKPGVYTRPLPLFRDWIKENTGV 683
Db 833 RNKPGVYTRPLPLFRDWIKENTGV 855

RESULT 14
AAB98500
ID AAB98500 standard; protein; 855 AA.
XX AAB98500;
AC AAB98500;
DT 03-AUG-2001 (first entry)
XX 03-AUG-2001 (first entry)
DE Human TADG-15.
XX Human; TADG-15; cytosolic; vaccine; ovarian tumour; cancer;
KW tumour antigen-derived gene 15; extracellular serine protease.
XX Homo sapiens.
OS Homo sapiens.
XX MO200129056-A1.
PN 26-APR-2001.
PD 26-APR-2001.
XX 26-APR-2001.
PF 20-OCT-2000; 2000MO-US029095.
XX 20-OCT-2000; 2000MO-US029095.
PR 20-OCT-1999; 99US-00421213.
XX 20-OCT-1999; 99US-00421213.
PA (UYAR-) UNIV ARKANSAS.
PI O'Brien TJ, Tanimoto H;
XX WPI; 2001-381031/40.
DR N-PSDB; AAH23601.
XX Novel extracellular serine protease, termed tumor antigen-derived gene 15
PT protein overexpressed in carcinomas and DNA encoding it, for diagnosis,
PT treatment, prevention of cancer, particularly breast, ovarian cancer.

XX Claim 11; Fig 2; 130pp; English.
PS The present sequence represents human tumour antigen-derived gene 15.
CC (TADG-15) protein. TADG-15 is an extracellular serine protease. It was
CC found that TADG-15 is over-expressed in ovarian tumours. TADG-15 protein
CC or its fragments of 9-40 residues that lack TADG-15 protease activity are
CC useful for vaccinating an individual against TADG-15, having, suspected
CC of having or at risk of getting cancer. Furthermore, the TADG-15 gene can
CC be used as a diagnostic or therapeutic target in cancer
XX

Sequence 855 AA;
SQ

Query Match 99.1%; Score 3754; DB 4; Length 855;
Best Local Similarity 99.3%; Pred. No. 1.2e-240;
Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

1 MAERVVMLPPRASLSFVTVSVAPPTDSKTVQRTQDNCSPGLHARGVELMRFTTPG 60
Db 173 MAERVVMLPPRASLSFVTVSVAPPTDSKTVQRTQDNCSPGLHARGVELMRFTTPG 232
Qy 61 PPDSPPYAHARQWALRGDADSVLSLTFRSFDLASCDERSGDLVTVYNTLSPMEPHALVQ 120
Db 233 PPDSPPYAHARQWALRGDADSVLSLTFRSFDLASCDERSGDLVTVYNTLSPMEPHALVQ 292
Qy 121 LCGTPPSYNTLTFHSSQVLLITLITNTERHHPGEATFFQLPKMSSCGRLRKAQGTFN 180
Db 293 LCGTPPSYNTLTFHSSQVLLITLITNTERHHPGEATFFQLPKMSSCGRLRKAQGTFN 352
Qy 181 SPYPGHYPNIDCTWNIIEVNNQHVKVRFPKPYLLBPRRACGTCPKDYVEINSEKTCGE 240
Db 353 SPYPGHYPNIDCTWNIIEVNNQHVKVSFPKPYLLBEGVAGTCPKDYVEINSEKTCGE 412
Qy 241 RSQFVTVSNKKTIVRFHSDSYDTGTGLAEYLSYDSDPCPGQFTCRGCRKELKCD 300
Db 413 RSQFVTVSNKKTIVRFHSDSYDTGTGLAEYLSYDSDPCPGQFTCRGCRKELKCD 472
Qy 301 GMACTDHSDELNCSDAHQFTCKNKECKPLFWVCDSDVNDGSDSGSCPAQTRFC 360
Db 473 GMACTDHSDELNCSDAHQFTCKNKECKPLFWVCDSDVNDGSDSGSCPAQTRFC 532
Qy 361 SNGKCLSKSQCGNKGDCGSDSDASCPKVVVVTCTKHTYRCLNGLCLSKNPECDGKED 420
Db 533 SNGKCLSKSQCGNKGDCGSDSDASCPKVVVVTCTKHTYRCLNGLCLSKNPECDGKED 592
Qy 421 CSDGSDXDCDCCGLRSFTQARVVGTDADGEMPMQVSLHALGQGHICGASLISPMWLV 480
Db 593 CSDGSDXDCDCCGLRSFTQARVVGTDADGEMPMQVSLHALGQGHICGASLISPMWLV 652
Qy 481 SAACIYIDDRGFRYSDFQWTAFLGLHDQSORSAFGVERLKRITISHPFNDFTFDYDI 540
Db 653 SAACIYIDDRGFRYSDFQWTAFLGLHDQSORSAFGVERLKRITISHPFNDFTFDYDI 712
Qy 541 ALLELEKPAEYSSMWRIICLPDASHVPAGKAIWVTGNGHTQYGGTALILQKEIRVIN 600
Db 713 ALLELEKPAEYSSMWRIICLPDASHVPAGKAIWVTGNGHTQYGGTALILQKEIRVIN 772
Qy 601 QTTCEMLLPQOITPRMNCVGLSGVDSGQSDSGPLSSVADGRIIFQAGVWSMGDCAQ 660
Db 773 QTTCEMLLPQOITPRMNCVGLSGVDSGQSDSGPLSSVADGRIIFQAGVWSMGDCAQ 832
Qy 661 RNKPGVYTRPLPLFRDWIKENTGV 683
Db 833 RNKPGVYTRPLPLFRDWIKENTGV 855

RESULT 15
AAB06930
ID AAB06930 standard; protein; 855 AA.
XX AAB06930;
AC AAB06930;
DT 16-OCT-2001 (first entry)

XX	Human membrane-type serine protease (MTSP) 1.
DE	Human; transmembrane serine protease; membrane-type serine protease;
XX	Human; transmembrane serine protease; membrane-type serine protease;
KM	lung carcinoma; colon adenocarcinoma; ovarian carcinoma; gene therapy;
KM	matrilysin.
XX	
OS	Homo sapiens.
XX	
FT	Key
FT	Domain
XX	Location/Qualifiers
XX	615..855
XX	/label= Protease_domain
XX	
XX	WO200157194-A2.
XX	
XX	09-AUG-2001.
XX	
XX	02-FEB-2001; 2001WO-US003471.
XX	
XX	03-FEB-2000; 2000US-0179962P.
XX	18-FEB-2000; 2000US-0183542P.
XX	22-JUN-2000; 2000US-0213124P.
XX	26-JUL-2000; 2000US-0220970P.
XX	08-SEP-2000; 2000US-00657986.
XX	22-SEP-2000; 2000US-0234840P.
XX	
XX	(CORV-) CORVAS INT INC.
XX	
PI	Madison EL, Ong EO, Yeh J;
XX	
XX	WPI; 2001-488877/53.
XX	N-PSDB; AAD13113.
XX	
PT	Novel single chain polypeptide comprising protease domain of type-II
PT	membrane-type serine protease or its catalytically active portion useful
PT	for treating and preventing cancer and tumor.
XX	
XX	Claim 12; Page 195-197; 256pp; English.
XX	
XX	The invention relates to transmembrane serine proteases and their
XX	corresponding nucleotides and the protease domain of a type-II membrane-
XX	type serine protease (MTSP). MTSP is useful for identifying compounds
XX	that modulate or inhibit its proteolytic activity and for formulating a
XX	medicament for treating neoplastic disease. MTSP and its corresponding
XX	nucleotides are useful in preventing or treating tumors or cancers such
XX	as lung carcinoma, colon adenocarcinoma and ovarian carcinoma, in
XX	diagnostics and in hybridisation assays. MTSP is useful as a diagnostic
XX	marker for tumour development, growth and/or progression and as
XX	immunogens to generate antibodies that specifically bind to it. MTSP DNA
XX	is useful in a yeast two-hybrid system and in gene therapy. The present
XX	sequence is human MTSP1 protein (also called matrilysin)
XX	
XX	Sequence 855 AA;
XX	
XX	Query Match 99.1%; Score 3754; DB 4; Length 855;
XX	Best local Similarity 99.3%; Pred. No. 1.2e-240;
XX	Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
XX	
QY	1 MAERVVMLPPRASLSFVVTSTVAAPPTDSKTYQRTQDNSCGSLGARGVELMRFPTTPG 60
DB	173 MAERVVMLPPRASLSFVVTSTVAAPPTDSKTYQRTQDNSCGSLGARGVELMRFPTTPG 232
QY	61 FPDSPYPAHARCOALRGDADSVSLTFRSPDLASCBERGSDLVTVVNTLSPMEPHALVQ 120
DB	233 FPDSPYPAHARCOALRGDADSVSLTFRSPDLASCBERGSDLVTVVNTLSPMEPHALVQ 292
QY	121 LCGTPSPSYNLTFHSSONVLLITLITNTERRHGFEATFFQLPRMSSCGGLRAAQGTFN 180
DB	293 LCGTPSPSYNLTFHSSONVLLITLITNTERRHGFEATFFQLPRMSSCGGLRAAQGTFN 352
QY	181 SPYRGHVPPIIDCTWNIIEVNNQHVKKRFFFLLEPRRACGTCPCDYVINEKTCGE 240

DB	353 SPYRGHVPPIIDCTWNIIEVNNQHVKKRFFFLLEPRRACGTCPCDYVINEKTCGE 412
QY	241 RSQFVTSNKNKITTFRHSDQSYTDGTFLAEYLSYDSDPCPQFTGRTGRTKRELRCD 300
DB	413 RSQFVTSNKNKITTFRHSDQSYTDGTFLAEYLSYDSDPCPQFTGRTGRTKRELRCD 472
QY	301 GMADCTDHSDELNCSCDAGHQFTCKNFKCPFLFWVCDSVNDCGNSDEGCSCPAQTFR 360
DB	473 GMADCTDHSDELNCSCDAGHQFTCKNFKCPFLFWVCDSVNDCGNSDEGCSCPAQTFR 532
QY	361 SNGKCLSKSQOQCKGKDDCGSDPASCPRNVVTCNHTYRCLNGLCLSGNPECDKED 420
DB	533 SNGKCLSKSQOQCKGKDDCGSDPASCPRNVVTCNHTYRCLNGLCLSGNPECDKED 592
QY	421 CSDGSEKDCDCGLRSTFRQARVVGTDADGEMPMQVSLHALGQGHICGASLSPNMLV 480
DB	593 CSDGSEKDCDCGLRSTFRQARVVGTDADGEMPMQVSLHALGQGHICGASLSPNMLV 652
QY	481 SAHCYTDGRGFRYSDEPTQMTAFLGLHDQSRSAFVQERRLKRIISHPFNDFTPDYDI 540
DB	653 SAHCYTDGRGFRYSDEPTQMTAFLGLHDQSRSAFVQERRLKRIISHPFNDFTPDYDI 712
QY	541 ALLELEKPAEYSMWPRICLPDASHVPKAKAIWVGWGHTOYGGTGALILQGEIRVIN 600
DB	713 ALLELEKPAEYSMWPRICLPDASHVPKAKAIWVGWGHTOYGGTGALILQGEIRVIN 772
QY	601 QTTCENTLLPQOITPRMNCVGLSGVDSCGDSGSPLSVEADGRIFFQAGVSWGDCQAQ 660
DB	773 QTTCENTLLPQOITPRMNCVGLSGVDSCGDSGSPLSVEADGRIFFQAGVSWGDCQAQ 832
QY	661 RNKRGVYTRLLPLFRDWIKENTGV 683
DB	833 RNKRGVYTRLLPLFRDWIKENTGV 855
XX	
XX	RESULT 16
XX	AAO22929
XX	ID AAO22929 standard; protein; 855 AA.
XX	
XX	AAO22929;
XX	
XX	12-DEC-2002 (first entry)
XX	
XX	Type II transmembrane serine protease 1 protein SEQ ID NO 2.
XX	
XX	Cytostatic; type-II membrane-type serine protease 7; MTSP7; malignancy;
XX	neoplastic disease; pre-malignant lesion; tumour; zymogen form; cancer;
XX	malignant; enzyme.
XX	
XX	Homo sapiens.
XX	
XX	WO200272786-A2.
XX	
XX	19-SEP-2002.
XX	
XX	13-MAR-2002; 2002WO-US007903.
XX	
XX	13-MAR-2001; 2001US-0275592P.
XX	
XX	(CORV-) CORVAS INT INC.
XX	
XX	Madison EL, Ong EO;
XX	
XX	WPI; 2002-732827/79.
XX	N-PSDB; AAL53444.
XX	
XX	New transmembrane serine protease 7 (MTSP7) polypeptide for diagnosing
XX	neoplastic diseases, monitoring tumor progress or therapeutic
XX	effectiveness, or identifying MTSP7 modulators for treating tumors or
XX	cancers.
XX	
XX	Disclosure; Page 172-174; 184pp; English.

CC The invention relates to a purified single or two-chain polypeptide,
CC which comprises the protease domain of a type-II membrane-type serine
CC protease 7 (MSP7) or its catalytically active portion. The polypeptide
CC comprising MSP7 is useful for detecting or diagnosing a neoplastic
CC disease, a pre-malignant lesion, a malignancy or other pathologic
CC condition in a subject. This polypeptide is also useful for monitoring
CC tumour (e.g. tumour of the breast, cervix, prostate, lung, ovary or
CC colon) progress and/or therapeutic effectiveness. The inhibitor of the
CC polypeptide or the inhibitor of the polypeptide's zymogen form is useful
CC for treating or preventing a neoplastic disease, or tumour initiation,
CC growth or progression, or a (pre-)malignant condition. The polypeptide or
CC polynucleotide is also useful for identifying modulators of MSP7, which
CC may be used to treat cancers or tumours. This sequence represents a
CC protein of the type-II membrane-type serine protease 1 relating to the
CC invention

XX
XX
SQ Sequence 855 AA;

Query Match 99.1%; Score 3754; DB 5; Length 855;

Best Local Similarity 99.3%; Pred. No. 1.2e-240; Mismatches 0; Gaps 0;

Matches 678; Conservative 0; Indels 0; Gaps 0;

QY 1 MAERVVMLEPPRATSLKSFVTVSVAFPTDSKTVQRTQDSCSFGIARGVELARFTTPG 60
DB 173 MAERVVMLEPPRATSLKSFVTVSVAFPTDSKTVQRTQDSCSFGIARGVELARFTTPG 232
QY 61 PPDSPPYAHARCGVALRGDADSVLSLTRSPDLASCDERGDLYTVYNTLSPEMHALVQ 120
DB 233 PPDSPPYAHARCGVALRGDADSVLSLTRSPDLASCDERGDLYTVYNTLSPEMHALVQ 292
QY 121 LCGYPPSYNLTFFSSQNVLLITLTINERHPGEATFQPLPMSSCGRLRAQGFEN 180
DB 293 LCGYPPSYNLTFFSSQNVLLITLTINERHPGEATFQPLPMSSCGRLRAQGFEN 352
QY 181 SPYYPGHYPNIDCTWNIENPNQHVKKVFEFYLLBERRACGTCPKDYVEINEKCYGE 240
DB 353 SPYYPGHYPNIDCTWNIENPNQHVKKVFEFYLLBERRACGTCPKDYVEINEKCYGE 412
QY 241 RSQGVVTSNSKIVTRFHSDSYTDGFLAFLYSYSSDPCEGQFTCTGHCIRKELRCD 300
DB 413 RSQGVVTSNSKIVTRFHSDSYTDGFLAFLYSYSSDPCEGQFTCTGHCIRKELRCD 472
QY 301 GMACTHSDLELNCSCAGHOFCTCKNFKCKPLFWVCDSVNDGNSDQSCSPAOTFRC 360
DB 473 GMACTHSDLELNCSCAGHOFCTCKNFKCKPLFWVCDSVNDGNSDQSCSPAOTFRC 532
QY 361 SNGKCLSKSQCGNKGDCGDSDBASCPCXNVVVTCTKATYCLNGLCLSKGNPECDGED 420
DB 533 SNGKCLSKSQCGNKGDCGDSDBASCPCXNVVVTCTKATYCLNGLCLSKGNPECDGED 592
QY 421 CSDGSDKDCDCCGLRSFTROARVVGGTDADGEMPMOVSLLHALCGHICGASLISPMWLV 480
DB 593 CSDGSDKDCDCCGLRSFTROARVVGGTDADGEMPMOVSLLHALCGHICGASLISPMWLV 652
QY 481 SAAHCYIDDGFRYSDDPTQWTAFLGLHDQSORSAFCVGERLKRLLISHPFNDDTFPVDI 540
DB 653 SAAHCYIDDGFRYSDDPTQWTAFLGLHDQSORSAFCVGERLKRLLISHPFNDDTFPVDI 712
QY 541 ALLELEKPAEYSSAWRPICLPDASHVPFAGKAIWYTGHTQYGGTGLLIQKEIRIYN 600
DB 713 ALLELEKPAEYSSAWRPICLPDASHVPFAGKAIWYTGHTQYGGTGLLIQKEIRIYN 772
QY 601 QTTCENTLLPQOITRRMNCVGLSGGVDSQCGDSGGPLSSVEADGRIFQAGVSWGDCAQ 660
DB 773 QTTCENTLLPQOITRRMNCVGLSGGVDSQCGDSGGPLSSVEADGRIFQAGVSWGDCAQ 832
QY 661 RNKPGVYTRLPFLFRDWIKENTGV 683
DB 833 RNKPGVYTRLPFLFRDWIKENTGV 855

RESULT 17
AD116816

ID AD116816 standard; protein; 855 AA.
XX
AC AD116816;
XX
DT 15-APR-2004 (first entry)
XX
DE Human NOVX protein homologue Segid 352.
XX
KW human; NOVX; cardiomyopathy; atherosclerosis; cancer; diabetes;
KW inflammation; autoimmune disorder; allergy; blood disorder;
KW acquired immunodeficiency syndrome; AIDS; obesity; asthma;
KW immunoglobulin (Ig)A nephropathy; cirrhosis; arthritis;
KW Alzheimer's disease; infection; str.
XX
OS Homo sapiens.
XX
PN MO20026849-A2.
XX
PD 06-SEP-2002.
XX
PP 31-JAN-2002; 2002MO-US002785.
XX
PR 31-JAN-2001; 2001US-0265395P.
PR 31-JAN-2001; 2001US-0265412P.
PR 31-JAN-2001; 2001US-0265514P.
PR 31-JAN-2001; 2001US-0265517P.
PR 02-FEB-2001; 2001US-0266406P.
PR 05-FEB-2001; 2001US-0266767P.
PR 07-FEB-2001; 2001US-0266975P.
PR 07-FEB-2001; 2001US-0267057P.
PR 08-FEB-2001; 2001US-0267459P.
PR 09-FEB-2001; 2001US-0267823P.
PR 15-FEB-2001; 2001US-0268974P.
PR 26-FEB-2001; 2001US-0271664P.
PR 27-FEB-2001; 2001US-0271839P.
PR 27-FEB-2001; 2001US-0271855P.
PR 02-MAR-2001; 2001US-0272788P.
PR 02-MAR-2001; 2001US-0273046P.
PR 14-MAR-2001; 2001US-0275925P.
PR 14-MAR-2001; 2001US-0275947P.
PR 14-MAR-2001; 2001US-0275950P.
PR 14-MAR-2001; 2001US-0275989P.
PR 15-MAR-2001; 2001US-0276448P.
PR 15-MAR-2001; 2001US-0276450P.
PR 16-MAR-2001; 2001US-0276377P.
PR 16-MAR-2001; 2001US-0276768P.
PR 20-MAR-2001; 2001US-0278652P.
PR 26-MAR-2001; 2001US-0278775P.
PR 26-MAR-2001; 2001US-0278778P.
PR 29-MAR-2001; 2001US-0279882P.
PR 29-MAR-2001; 2001US-0279884P.
PR 30-MAR-2001; 2001US-0280147P.
PR 11-APR-2001; 2001US-0282992P.
PR 11-APR-2001; 2001US-0283083P.
PR 20-APR-2001; 2001US-0285133P.
PR 23-APR-2001; 2001US-0285749P.
PR 03-MAY-2001; 2001US-0288377P.
PR 03-MAY-2001; 2001US-0288504P.
PR 29-MAY-2001; 2001US-0294047P.
PR 30-MAY-2001; 2001US-0294473P.
PR 08-JUN-2001; 2001US-0296964P.
PR 18-JUN-2001; 2001US-0298959P.
PR 19-JUN-2001; 2001US-0299324P.
PR 13-AUG-2001; 2001US-0312020P.
PR 16-AUG-2001; 2001US-0312889P.
PR 16-AUG-2001; 2001US-0312908P.
PR 21-AUG-2001; 2001US-0313390P.
PR 28-AUG-2001; 2001US-0315470P.
PR 31-AUG-2001; 2001US-0316447P.
PR 07-SEP-2001; 2001US-0318115P.
PR 07-SEP-2001; 2001US-0318118P.
PR 12-SEP-2001; 2001US-0318740P.
PR 19-SEP-2001; 2001US-0323379P.

PR 18-OCT-2001; 2001US-0330245P.
 PR 18-OCT-2001; 2001US-0330308P.
 PR 14-NOV-2001; 2001US-0332701P.
 XX
 PA (CURA-) CURAGEN CORP.
 PI Therneer VT, Spytek KA, Zernusen BD, Paturajan M, Shinkens RA;
 PI Li L, Ganggoli EA, Radigan M, Anderson DW, Rastelli L, Miller CE;
 PI Gerlach VL, Taupier RJ, Gause VY, Colman SD, Wolenc AR, Pena CEA;
 PI Futrak K, Grose WM, Alsbrook JP, Lepley DM, Rieger DK, Burgess CE;
 XX
 DR WPI; 2002-706998/76.
 XX
 XX New NOVX polypeptides and nucleic acids, useful for preventing or
 PT treating NOVX-associated disorders, e.g. cancer, cardiomyopathy,
 PT atherosclerosis, or diabetes, and in chromosome mapping, tissue typing or
 PT pharmacogenomics.
 XX
 PS Disclosure, SEQ ID NO 352; 1498bp; English.
 XX
 XX This invention relates to a novel nucleic acids, and encoded polypeptides
 CC thereof, which have properties related to the stimulation of biochemical
 CC or physiological responses in a cell, tissue, organ or organism.
 CC Specifically, it refers to the use of biologically active fragments for
 CC diagnostic and prognostic assays and furthermore in the treatment of
 CC diverse pathological conditions. The present invention describes novel
 CC human and murine NOVX proteins, as well as methods to modulate their
 CC expression using antisense oligos, ribozymes and peptide nucleic acids.
 CC The NOVX polypeptides, polynucleotides and antibodies are useful in
 CC treating or preventing NOVX-associated disorders, e.g. cardiomyopathy,
 CC atherosclerosis, cancer and diabetes. Furthermore, they may be used in
 CC treating or preventing diseases such as inflammation, autoimmune
 CC disorders, allergies, blood disorders, acquired immunodeficiency syndrome
 CC (AIDS), obesity, asthma, immunoglobulin (Ig)A nephropathy, cirrhosis,
 CC arthritis, Alzheimer's disease, infections, stroke, muscular dystrophy
 CC and epilepsy. Accordingly, these molecules have many activities including
 CC cytoskeletal, cardiac, antiinflammatory, immunosuppressive, antiallergic,
 CC hemostatic, anti-HIV, antidiabetic, antiatherosclerotic, anorectic,
 CC antistatic, nephrotropic, antiautistic, hepatotropic,
 CC neuroprotective, nootropic, antibacterial, virucide, antiparasitic,
 CC relaxant and anticonvulsant. In addition, they are useful in screening
 CC assays to identify small molecules that modulate or inhibit, for example,
 CC neurogenesis, wound healing and angiogenesis. The nucleic acids are also
 CC used as in chromosome mapping, tissue typing, preventive medicine and
 CC pharmacogenomics. This polypeptide is a homologue of a human NOVX protein
 CC of the invention.
 CC
 CC
 XX
 XX Sequence 855 AA:
 SQ

Query Match 99.1%; Score 3754; DB 5; Length 855;
 Best Local Similarity 99.3%; Pred. No. 1.2e-240;
 Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAEERVVMLPPRAARKSFVTVSVAPPTSKTYQRTODNSCGRGHARGVEIMRFTTPG 60
 DB 173 MAEERVVMLPPRAARKSFVTVSVAPPTSKTYQRTODNSCGRGHARGVEIMRFTTPG 232
 QY 61 PPDSPYAHARCGWALRGDADSVLSTFRSFDLASCDERSGDLVTVNTLSPMEPHALVQ 120
 DB 233 PPDSPYAHARCGWALRGDADSVLSTFRSFDLASCDERSGDLVTVNTLSPMEPHALVQ 232
 QY 121 LCGTTPSYNLTFHSSQNVLLITLITNTERRHHPGFATFQLPMSSCGRLRKAQGTEN 180
 DB 293 LCGTTPSYNLTFHSSQNVLLITLITNTERRHHPGFATFQLPMSSCGRLRKAQGTEN 352
 QY 181 SPVYPGHYPNIDCTNMIIEVNNQHVVRKFFYLEPRACGCPDYVINEKXCGE 240
 DB 353 SPVYPGHYPNIDCTNMIIEVNNQHVVRKFFYLEPRACGCPDYVINEKXCGE 412
 QY 241 RSQFVVVSNENKLTVRHSDQSYDTDTGFLAEVLSYSDSDPCPGQFTCRTRCIRKELRCD 300
 DB 413 RSQFVVVSNENKLTVRHSDQSYDTDTGFLAEVLSYSDSDPCPGQFTCRTRCIRKELRCD 472

QY 301 GMADCTDHSDELNCSGDAGHQFTCKRKFCPLFWVCDSVNDCGDNDEGCGSCPACTQTRFC 360
 DB 473 GMADCTDHSDELNCSGDAGHQFTCKRKFCPLFWVCDSVNDCGDNDEGCGSCPACTQTRFC 532
 QY 361 SNKGKLSKSGQCGKXDCDGDGSDGDEASCPRKNNVYTCFKHYTRCLANGCLSKGNECGKED 420
 DB 533 SNKGKLSKSGQCGKXDCDGDGSDGDEASCPRKNNVYTCFKHYTRCLANGCLSKGNECGKED 532
 QY 421 CSDGSDKDCDCGLRSFTQARVVGCTDADGEMPMQVSLHAGQGHICGASLISPNWLV 480
 DB 593 CSDGSDKDCDCGLRSFTQARVVGCTDADGEMPMQVSLHAGQGHICGASLISPNWLV 652
 QY 481 SAHACTIDRGFRFYSPTQMTAVLGLHDQSQRABEVORRLKRITISHPFNDFEDYI 540
 DB 653 SAHACTIDRGFRFYSPTQMTAVLGLHDQSQRABEVORRLKRITISHPFNDFEDYI 712
 QY 541 ALLEKPAEYSMWBPICLPDASHVFPAGKAIWVGWGHQYGTGALILQKEIRVIN 600
 DB 713 ALLEKPAEYSMWBPICLPDASHVFPAGKAIWVGWGHQYGTGALILQKEIRVIN 772
 QY 601 QTTCEMLPQOITPRMNCVGLSGVDSQCGDSGGLSSVEADGRIFOAGVSWGDCAQ 660
 DB 773 QTTCEMLPQOITPRMNCVGLSGVDSQCGDSGGLSSVEADGRIFOAGVSWGDCAQ 832
 QY 661 RNKPGVYTRPLFRDWIKENTGV 683
 DB 833 RNKPGVYTRPLFRDWIKENTGV 855

RESULT 18
 AD116882
 ID AD116882 standard; protein; 855 AA.
 XX
 XX AD116882;
 AC
 XX
 DT 15-APR-2004 (first entry)
 XX
 XX Human NOVX protein homologue SegID 418.
 DE
 XX human; NOVX; cardiomyopathy; atherosclerosis; cancer; diabetes;
 KW inflammation; autoimmune disorder; allergy; blood disorder;
 KW acquired immunodeficiency syndrome; AIDS; obesity; asthma;
 KW immunoglobulin (Ig)A nephropathy; cirrhosis; arthritis;
 KW Alzheimer's disease; infection; str.
 KW
 XX
 OS Homo sapiens.
 OS
 XX
 PN WO200268649-A2.
 PN
 XX
 PD 06-SEP-2002.
 PD
 XX
 PF 31-JAN-2002; 2002WO-US002785.
 PF
 XX
 PR 31-JAN-2001; 2001US-0265395P.
 PR 31-JAN-2001; 2001US-0265412P.
 PR 31-JAN-2001; 2001US-0265514P.
 PR 31-JAN-2001; 2001US-0265517P.
 PR 02-FEB-2001; 2001US-0266406P.
 PR 05-FEB-2001; 2001US-0266767P.
 PR 07-FEB-2001; 2001US-0266975P.
 PR 07-FEB-2001; 2001US-0267057P.
 PR 08-FEB-2001; 2001US-0267459P.
 PR 09-FEB-2001; 2001US-0267823P.
 PR 15-FEB-2001; 2001US-0268974P.
 PR 26-FEB-2001; 2001US-0271664P.
 PR 27-FEB-2001; 2001US-0271839P.
 PR 27-FEB-2001; 2001US-0271855P.
 PR 02-MAR-2001; 2001US-0272788P.
 PR 02-MAR-2001; 2001US-0273046P.
 PR 14-MAR-2001; 2001US-0275925P.
 PR 14-MAR-2001; 2001US-0275947P.
 PR 14-MAR-2001; 2001US-0275950P.
 PR 14-MAR-2001; 2001US-0275989P.

PR 15-MAR-2001; 2001US-0276448P.
 PR 15-MAR-2001; 2001US-0276450P.
 PR 16-MAR-2001; 2001US-0276397P.
 PR 16-MAR-2001; 2001US-0276768P.
 PR 20-MAR-2001; 2001US-0278652P.
 PR 26-MAR-2001; 2001US-0278775P.
 PR 26-MAR-2001; 2001US-0278778P.
 PR 29-MAR-2001; 2001US-0279882P.
 PR 29-MAR-2001; 2001US-0279884P.
 PR 30-MAR-2001; 2001US-0280147P.
 PR 11-APR-2001; 2001US-0282992P.
 PR 11-APR-2001; 2001US-0283083P.
 PR 20-APR-2001; 2001US-0285133P.
 PR 23-APR-2001; 2001US-0285749P.
 PR 03-MAY-2001; 2001US-0288327P.
 PR 03-MAY-2001; 2001US-0288504P.
 PR 29-MAY-2001; 2001US-0294047P.
 PR 30-MAY-2001; 2001US-0294473P.
 PR 08-JUN-2001; 2001US-0296964P.
 PR 18-JUN-2001; 2001US-0296959P.
 PR 19-JUN-2001; 2001US-0299324P.
 PR 13-AUG-2001; 2001US-0312020P.
 PR 16-AUG-2001; 2001US-0312889P.
 PR 16-AUG-2001; 2001US-0312908P.
 PR 21-AUG-2001; 2001US-0313390P.
 PR 28-AUG-2001; 2001US-0315470P.
 PR 31-AUG-2001; 2001US-0316447P.
 PR 07-SEP-2001; 2001US-0318115P.
 PR 07-SEP-2001; 2001US-0318118P.
 PR 12-SEP-2001; 2001US-0318740P.
 PR 19-SEP-2001; 2001US-0323379P.
 PR 18-OCT-2001; 2001US-0330245P.
 PR 18-OCT-2001; 2001US-0330308P.
 PR 14-NOV-2001; 2001US-0332701P.

XX (CURA-) CURAGEN CORP.

PI Tchernev VT, Splytek KA, Zernusen BD, Patturajan M, Shinketa RA;
 PI Li L, Gangoli EA, Padigara M, Anderson DM, Rastelli L, Miller CE;
 PI Gerlach VL, Taupier R, Gusev VY, Colman SD, Wolenc AR, Pena CEA,
 PI Futak K, Grosse WM, Alsbrook JP, Lepley DM, Rieger DK, Burgess CE;
 XX WPI; 2002-706998/76.

PT New NOVX polypeptides and nucleic acids, useful for preventing or
 PT treating NOVX-associated disorders, e.g. cancer, cardiomyopathy,
 PT atherosclerosis, or diabetes, and in chromosome mapping, tissue typing or
 PT pharmacogenomics.

PS Disclosure; SEQ ID NO 418; 1498bp; English.

XX This invention relates to a novel nucleic acid, and encoded polypeptides
 CC thereof, which have properties related to the stimulation of biochemical
 CC or physiological responses in a cell, tissue, organ or organism.
 CC Specifically, it refers to the use of biologically active fragments for
 CC diagnostic and prognostic assays and furthermore in the treatment of
 CC diverse pathological conditions. The present invention describes novel
 CC human and murine NOVX proteins, as well as methods to modulate their
 CC expression using antisense oligos, ribozymes and peptide nucleic acids.
 CC The NOVX polypeptides, polynucleotides and antibodies are useful in
 CC treating or preventing NOVX-associated disorders, e.g. cardiomyopathy,
 CC atherosclerosis, cancer and diabetes. Furthermore, they may be used in
 CC treating or preventing diseases such as inflammation, autoimmune
 CC disorders, allergies, blood disorders, acquired immunodeficiency syndrome
 CC (AIDS), obesity, asthma, immunoglobulin (Ig) A nephropathy, cirrhosis,
 CC arthritis, Alzheimer's disease, infections, stroke, muscular dystrophy
 CC and epilepsy. Accordingly, these molecules have many activities including
 CC cytoskeletal, cardiac, antiinflammatory, immunosuppressive, antiallergic,
 CC haemostatic, anti-HIV, antidiabetic, antiarteriosclerotic, anorectic,
 CC antiasthmatic, nephroprotective, antiarthritic, hepatocytic,
 CC neuroprotective, neurotropic, antibacterial, vincine, antiparasitic,
 CC relaxant and anticonvulsant. In addition, they are useful in screening
 CC assays to identify small molecules that modulate or inhibit, for example,

CC neurogenesis, wound healing and angiogenesis. The nucleic acids are also
 CC used as in chromosome mapping, tissue typing, preventive medicine and
 CC pharmacogenomics. This polypeptide is a homologue of a human NOVX protein
 CC of the invention.

XX Sequence 855 AA;

Query Match 99.1%; Score 3754; DB 5; Length 855;
 Best Local Similarity 99.3%; Pred. No. 1.2e-240;
 Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAERVVMLEPPRARSLKSPVTVSVAPETDSTKVQRTQDNCSCFGLHARGVELMRFTTPG 60
 DB MAERVVMLEPPRARSLKSPVTVSVAPETDSTKVQRTQDNCSCFGLHARGVELMRFTTPG 232
 QY 61 PPDSPYPAHARCCQALRGDADSVLSLTFRSPDLASCDBERSDLYVTYNTLSPMEPALVQ 120
 DB PPDSPYPAHARCCQALRGDADSVLSLTFRSPDLASCDBERSDLYVTYNTLSPMEPALVQ 292
 QY 121 LCGTYPSPYNTLTFHSSQNVLLITLTNTERRHGFEPFQLPMSCGRLRAQGTFN 180
 DB LCGTYPSPYNTLTFHSSQNVLLITLTNTERRHGFEPFQLPMSCGRLRAQGTFN 352
 QY 293 LCGTYPSPYNTLTFHSSQNVLLITLTNTERRHGFEPFQLPMSCGRLRAQGTFN 352
 DB LCGTYPSPYNTLTFHSSQNVLLITLTNTERRHGFEPFQLPMSCGRLRAQGTFN 412
 QY 181 SPYYPGHYPNNIDCTWNIIEVPNNQHVVRKFFYLLEPRRACGTCPKDYVEINGEKYCGE 240
 DB SPYYPGHYPNNIDCTWNIIEVPNNQHVVRKFFYLLEPRRACGTCPKDYVEINGEKYCGE 412
 QY 353 SPYYPGHYPNNIDCTWNIIEVPNNQHVVRKFFYLLEPRRACGTCPKDYVEINGEKYCGE 412
 DB SPYYPGHYPNNIDCTWNIIEVPNNQHVVRKFFYLLEPRRACGTCPKDYVEINGEKYCGE 472
 QY 241 RSQFVTVTSNKNKTVTRHSDQSYTDGTFLAEYLSYDSDPCPQFTRTGRCIRKELRCD 300
 DB RSQFVTVTSNKNKTVTRHSDQSYTDGTFLAEYLSYDSDPCPQFTRTGRCIRKELRCD 472
 QY 413 RSQFVTVTSNKNKTVTRHSDQSYTDGTFLAEYLSYDSDPCPQFTRTGRCIRKELRCD 472
 DB RSQFVTVTSNKNKTVTRHSDQSYTDGTFLAEYLSYDSDPCPQFTRTGRCIRKELRCD 532
 QY 301 GMADCTHSDBLNCSCDAGHOFCKNFKCKPLFWVCDSVNDCCGNSDEGSCSCPAQTFRC 360
 DB GMADCTHSDBLNCSCDAGHOFCKNFKCKPLFWVCDSVNDCCGNSDEGSCSCPAQTFRC 532
 QY 473 GMADCTHSDBLNCSCDAGHOFCKNFKCKPLFWVCDSVNDCCGNSDEGSCSCPAQTFRC 532
 DB GMADCTHSDBLNCSCDAGHOFCKNFKCKPLFWVCDSVNDCCGNSDEGSCSCPAQTFRC 592
 QY 361 SNGKCLSKSQQCNCKDCCGDSDEASCCKXNVVYCTKHTYRCNLGLCLSKGNPECCKED 420
 DB SNGKCLSKSQQCNCKDCCGDSDEASCCKXNVVYCTKHTYRCNLGLCLSKGNPECCKED 592
 QY 533 SNGKCLSKSQQCNCKDCCGDSDEASCCKXNVVYCTKHTYRCNLGLCLSKGNPECCKED 592
 DB SNGKCLSKSQQCNCKDCCGDSDEASCCKXNVVYCTKHTYRCNLGLCLSKGNPECCKED 652
 QY 421 CSDGSDKDCDCGTRSTRTRQARVVGCTDADBGEMPMQVSLHALGGQHCASLSPMWLV 480
 DB CSDGSDKDCDCGTRSTRTRQARVVGCTDADBGEMPMQVSLHALGGQHCASLSPMWLV 652
 QY 593 CSDGSDKDCDCGTRSTRTRQARVVGCTDADBGEMPMQVSLHALGGQHCASLSPMWLV 652
 DB CSDGSDKDCDCGTRSTRTRQARVVGCTDADBGEMPMQVSLHALGGQHCASLSPMWLV 712
 QY 481 SAACHYIDDRGFRSPDPTQMTAFGLHDOSRSPAGVBERLRKRITISHPFNPTFFYDI 540
 DB SAACHYIDDRGFRSPDPTQMTAFGLHDOSRSPAGVBERLRKRITISHPFNPTFFYDI 712
 QY 653 SAACHYIDDRGFRSPDPTQMTAFGLHDOSRSPAGVBERLRKRITISHPFNPTFFYDI 712
 DB SAACHYIDDRGFRSPDPTQMTAFGLHDOSRSPAGVBERLRKRITISHPFNPTFFYDI 772
 QY 541 ALLELEKPAEYSSNVRPICLPDASHVFPACKAIWVTGHTQVGGTGALLILOGEIRVIN 600
 DB ALLELEKPAEYSSNVRPICLPDASHVFPACKAIWVTGHTQVGGTGALLILOGEIRVIN 772
 QY 713 ALLELEKPAEYSSNVRPICLPDASHVFPACKAIWVTGHTQVGGTGALLILOGEIRVIN 772
 DB ALLELEKPAEYSSNVRPICLPDASHVFPACKAIWVTGHTQVGGTGALLILOGEIRVIN 832
 QY 601 QTTCENTLPPQITPRMNCVGFSLSGVDSCGDSGCPSSVEADGRITQAGVSWGDCAQ 660
 DB QTTCENTLPPQITPRMNCVGFSLSGVDSCGDSGCPSSVEADGRITQAGVSWGDCAQ 832
 QY 773 QTTCENTLPPQITPRMNCVGFSLSGVDSCGDSGCPSSVEADGRITQAGVSWGDCAQ 832
 DB QTTCENTLPPQITPRMNCVGFSLSGVDSCGDSGCPSSVEADGRITQAGVSWGDCAQ 852
 QY 661 RNKPGVYTRLPFPDWTKENTGV 683
 DB RNKPGVYTRLPFPDWTKENTGV 855
 QY 833 RNKPGVYTRLPFPDWTKENTGV 855
 DB RNKPGVYTRLPFPDWTKENTGV 855

RESULT 19

ID AD116875
 ID AD116875 standard; protein: 855 AA.

XX AD116875;

XX AD116875;

XX 15-Apr-2004 (first entry)

XX Human NOVX protein homologue SeqID 411.

XX human; NOVX; cardiomyopathy; atherosclerosis; cancer; diabetes;

KW inflammation; autoimmune disorder; allergy; blood disorder;
 KW acquired immunodeficiency syndrome; AIDS; obesity; asthma;

KW immunoglobulin (Ig)A nephropathy; cirrhosis; arthritis;
XX Alzheimer's disease; infection; str.
OS Homo sapiens.
XX WO200268649-A2.
XX
XX 06-SEP-2002.
XX
XX 31-JAN-2002; 2002MO-US002785.
XX
XX 31-JAN-2001; 2001US-0265395P.
XX 31-JAN-2001; 2001US-0265412P.
XX 31-JAN-2001; 2001US-0265514P.
XX 31-JAN-2001; 2001US-0265517P.
XX 02-FEB-2001; 2001US-0266406P.
XX 05-FEB-2001; 2001US-0266767P.
XX 07-FEB-2001; 2001US-0266975P.
XX 07-FEB-2001; 2001US-0267057P.
XX 08-FEB-2001; 2001US-0267459P.
XX 09-FEB-2001; 2001US-0267823P.
XX 15-FEB-2001; 2001US-0268974P.
XX 26-FEB-2001; 2001US-0271664P.
XX 27-FEB-2001; 2001US-0271839P.
XX 27-FEB-2001; 2001US-0271855P.
XX 02-MAR-2001; 2001US-0272788P.
XX 02-MAR-2001; 2001US-0273046P.
XX 14-MAR-2001; 2001US-0275925P.
XX 14-MAR-2001; 2001US-0275947P.
XX 14-MAR-2001; 2001US-0275950P.
XX 14-MAR-2001; 2001US-0275959P.
XX 15-MAR-2001; 2001US-0276448P.
XX 15-MAR-2001; 2001US-0276450P.
XX 16-MAR-2001; 2001US-0276397P.
XX 16-MAR-2001; 2001US-0276768P.
XX 20-MAR-2001; 2001US-0278652P.
XX 26-MAR-2001; 2001US-0278775P.
XX 26-MAR-2001; 2001US-0278778P.
XX 29-MAR-2001; 2001US-0279882P.
XX 29-MAR-2001; 2001US-0279884P.
XX 30-MAR-2001; 2001US-0280147P.
XX 11-APR-2001; 2001US-0282992P.
XX 11-APR-2001; 2001US-0283083P.
XX 20-APR-2001; 2001US-0285133P.
XX 23-APR-2001; 2001US-0285749P.
XX 03-MAY-2001; 2001US-0288327P.
XX 03-MAY-2001; 2001US-0288504P.
XX 29-MAY-2001; 2001US-0294047P.
XX 30-MAY-2001; 2001US-0294473P.
XX 08-JUN-2001; 2001US-0296964P.
XX 18-JUN-2001; 2001US-0298959P.
XX 19-JUN-2001; 2001US-0299324P.
XX 13-AUG-2001; 2001US-0312020P.
XX 16-AUG-2001; 2001US-0312889P.
XX 16-AUG-2001; 2001US-0312908P.
XX 21-AUG-2001; 2001US-0313390P.
XX 28-AUG-2001; 2001US-0315470P.
XX 31-AUG-2001; 2001US-0316447P.
XX 07-SEP-2001; 2001US-0318115P.
XX 07-SEP-2001; 2001US-0318118P.
XX 12-SEP-2001; 2001US-0318740P.
XX 19-SEP-2001; 2001US-0323379P.
XX 18-OCT-2001; 2001US-0330245P.
XX 18-OCT-2001; 2001US-0330308P.
XX 14-NOV-2001; 2001US-0332701P.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Tchernev VT, Spytek KA, Zethusen BD, Patturajan M, Shinkets RA;
PI Li L, Gangoli EA, Padigaru M, Anderson DW, Raselli L, Miller CE;
PI Gerlach VL, Taudier RJ, Gusev VY, Colman SD, Wolenc AR, Pena CE;
PI Pirtak K, Grose WM, Alsbrook JF, Lepley DM, Rieger DK, Burgess CE;

DR WPI, 2002-706998/76.
XX
XX New NOVX polypeptides and nucleic acids, useful for preventing or
PT treating NOVX-associated disorders, e.g. cancer, cardiomyopathy,
PT atherosclerosis, or diabetes, and in chromosome mapping, tissue typing or
PT pharmacogenomics.
XX
XX Disclosure; SEQ ID NO 411; 1498bp; English.
XX
XX This invention relates to a novel nucleic acids, and encoded polypeptides
CC thereof, which have properties related to the stimulation of biochemical
CC or physiological responses in a cell, tissue, organ or organism.
CC Specifically, it refers to the use of biologically active fragments for
CC diagnostic and prognostic assays and furthermore in the treatment of
CC diverse pathological conditions. The present invention describes novel
CC human and murine NOVX proteins, as well as methods to modulate their
CC expression using antisense oligos, ribozymes and peptide nucleic acids.
CC The NOVX polypeptides, polynucleotides and antibodies are useful in
CC treating or preventing NOVX-associated disorders, e.g. cardiomyopathy,
CC atherosclerosis, cancer and diabetes. Furthermore, they may be used in
CC treating or preventing diseases such as inflammation, autoimmune
CC disorders, allergies, blood disorders, acquired immunodeficiency syndrome
CC (AIDS), obesity, asthma, immunoglobulin (Ig)A nephropathy, cirrhosis,
CC arthritis, Alzheimer's disease, infections, stroke, muscular dystrophy,
CC and epilepsy. Accordingly, these molecules have many activities including
CC cytostatic, cardiant, antiinflammatory, immunosuppressive, antiallergic,
CC hemostatic, anti-HIV, antidiabetic, antiarteriosclerotic, anorectic,
CC antisthmatic, neurotropic, antiarthritic, hepatotropic,
CC neuroprotective, nootropic, antibacterial, viricide, antiparasitic,
CC relaxant and anticonvulsant. In addition, they are useful in screening
CC assays to identify small molecules that modulate or inhibit, for example,
CC neurogenesis, wound healing and angiogenesis. The nucleic acids are also
CC used as in chromosome mapping, tissue typing, preventive medicine and
CC pharmacogenomics. This polypeptide is a homologue of a human NOVX protein
CC of the invention.
XX
XX Sequence 855 AA;
SQ
Query Match 99.1%; Score 3754; DB 5; Length 855;
Best Local Similarity 99.3%; Pred. No. 1,2e-240;
Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 MAERVYMLPPRRSLKSVYTSVVAFPPTDSKVQORTQNSCFGHARGVEMRTTTC 60
DB 173 MAERVYMLPPRRSLKSVYTSVVAFPPTDSKVQORTQNSCFGHARGVEMRTTTPG 232
QY 61 FPDSPYPAHRCOMALRGDASVLSLTPRSFDLASDEKSDLVYVNTLSPMERPAIVQ 120
DB 233 FPDSPYPAHRCOMALRGDASVLSLTPRSFDLASDEKSDLVYVNTLSPMERPAIVQ 292
QY 121 LCGTYPSPYNLTFHSSQVLLITLITNTERRHGFEATFPQLEPRMSSCGRLRKAQGTEN 180
DB 293 LCGTYPSPYNLTFHSSQVLLITLITNTERRHGFEATFPQLEPRMSSCGRLRKAQGTEN 352
QY 181 SPYYPGHYPPNIDCTWNIIEVNNQHVKAFKFPYLLPERRACCTCKVDVEINGEYCCG 240
DB 353 SPYYPGHYPPNIDCTWNIIEVNNQHVKAFKFPYLLPERRACCTCKVDVEINGEYCCG 412
QY 241 RSQFVVTNSNKTITVFHSDQSYTDGFLAEYLSYSSPPCGQFCRTGRCIRKELRCD 300
DB 413 RSQFVVTNSNKTITVFHSDQSYTDGFLAEYLSYSSPPCGQFCRTGRCIRKELRCD 472
QY 301 GWADCTDHSDELNCSGDAGHQFTCKNFKFPLFWVCDVYNDGDNBDEGCGCPAQTFFC 360
DB 473 GWADCTDHSDELNCSGDAGHQFTCKNFKFPLFWVCDVYNDGDNBDEGCGCPAQTFFC 532
QY 361 SNKCKLSKSGQCKGKDCDGDSPDASCPKVNVTCTKHTYRCLNGCLSKGNPECCGKD 420
DB 533 SNKCKLSKSGQCKGKDCDGDSPDASCPKVNVTCTKHTYRCLNGCLSKGNPECCGKD 592
QY 421 CSDGSEKDCDGLRFTROARVVGSTDADEGEWPMQVSLHALGQSHIGASLISPNMTV 480
DB 593 CSDGSEKDCDGLRFTROARVVGSTDADEGEWPMQVSLHALGQSHIGASLISPNMTV 652

QY 481 SAACHYIDDRGFRYSDDPTQWTAFLGLHDQSORSAQVQERLKRILISHPFNDFTFYDI 540
 DB 653 SAACHYIDDRGFRYSDDPTQWTAFLGLHDQSORSAQVQERLKRILISHPFNDFTFYDI 712
 QY 541 ALLELEKPAEYSSWVRPICLPDASHVFPAGKAIWVTGMGHTQYGGTALILQKEIRVIN 600
 DB 713 ALLELEKPAEYSSWVRPICLPDASHVFPAGKAIWVTGMGHTQYGGTALILQKEIRVIN 772
 QY 601 QTTCEBNLLPQOITPRMNCVGFSLSGVDSCGDSGGPLSSVEADGRIFQAGVWSMGDCAQ 660
 DB 773 QTTCEBNLLPQOITPRMNCVGFSLSGVDSCGDSGGPLSSVEADGRIFQAGVWSMGDCAQ 832
 QY 661 RNKPGVYTRLPFRDWMKENTGV 683
 DB 833 RNKPGVYTRLPFRDWMKENTGV 855

RESULT 20
 ABP56619
 ID ABP56619 standard; protein: 855 AA.
 XX
 AC ABP56619;
 DT 24-MAR-2003 (first entry)
 XX
 DE Human membrane-type serine protease MTSP1 protein SEQ ID NO:2.
 XX
 KM Human; membrane-type serine protease; enzyme; MTSP10; cytosolic;
 KM type-II membrane-type serine protease; neoplastic disease; tumour; MTSP1;
 KM matrixase.
 XX
 OS Homo sapiens.
 XX
 PN WO200292841-A2.
 XX
 PD 21-NOV-2002.
 XX
 PF 14-MAY-2002; 2002W0-US015332.
 XX
 PR 14-MAY-2001; 2001US-0291001P.
 XX
 PA (CORV-) CORVAS INT INC.
 XX
 PI Madison BL, Yeh J;
 XX
 DR WPI: 2003-129309/12.
 DR N-PSDB; ABZ22450.
 XX
 PT New polypeptides comprising the protease domain of a type-II membrane-
 PT type serine protease (MTCP10), or its mutants, useful for diagnosing
 PT neoplasms or malignancies, or for screening for MTCP10 inhibitors for
 PT treating such diseases.
 XX
 PS Disclosure; Page 181-183; 198pp; English.
 XX
 CC The present invention describes a polypeptide comprising a purified
 CC single or two chain polypeptide, which comprises the protease domain of a
 CC type-II membrane-type serine protease (MTSP10) or its catalytically
 CC active portion, or a mutant of it, where up to 50 % of the amino acids
 CC are replaced with another amino acid, and the resulting polypeptide is a
 CC single chain or two chain polypeptide that has a catalytic activity of at
 CC least 1-10 % of the unmutated polypeptide. MTSP10 has cytosolic
 CC activity. The polypeptide containing the protease domain of the MTSP10 is
 CC useful for detecting a neoplastic disease, and for diagnosing the
 CC presence of a pre-malignant lesion, a malignancy, or other pathologic
 CC condition in a subject, or monitoring tumour (e.g. breast, cervix,
 CC prostate, lung, ovary or colon tumour) progression and/or therapeutic
 CC effectiveness. An inhibitor of the polypeptide containing the protease
 CC domain of MTSP10 is useful for treating or preventing neoplastic disease
 CC in a mammal. An inhibitor of the activation cleavage of the zymogen form
 CC of the MTSP10 polypeptide is useful for inhibiting tumour initiation,
 CC growth or progression, or treating (pre-)malignant conditions of the e.g.

CC breast, cervix, prostate, lung, ovary or colon. The present sequence
 CC represents human MTSP1 (also known as matrixase), which is used in an
 CC example from the present invention
 XX
 XX Sequence 855 AA;
 SQ
 Query Match 99.1%; Score 3754; DB 6; Length 855;
 Best Local Similarity 99.3%; Pred. No. 1,2e-240;
 Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 1 MAEERVVMLPPRARSLKSPVTVSVAPPTDSKTVOQTQDNCSCFGLHARGVELMRFTTPG 60
 DB 173 MAEERVVMLPPRARSLKSPVTVSVAPPTDSKTVOQTQDNCSCFGLHARGVELMRFTTPG 232
 QY 61 PPDSPYPAHARCCQALRGDADSVLSLTFRRSPDLASCBERGSDLVTVYNTLSPMEPALVQ 120
 DB 233 PPDSPYPAHARCCQALRGDADSVLSLTFRRSPDLASCBERGSDLVTVYNTLSPMEPALVQ 292
 QY 121 LCGTYPPSYNLTFHSSQNVLLITLITWTERRHQFEATFFQLPRMSSCGRLRKAOQTFN 180
 DB 293 LCGTYPPSYNLTFHSSQNVLLITLITWTERRHQFEATFFQLPRMSSCGRLRKAOQTFN 352
 QY 181 SPYYPGHYPPNIDCTWNIIEVPPNQHVVRKFFYLLEPRRACGTCPRDYVEINGEKYCGE 240
 DB 353 SPYYPGHYPPNIDCTWNIIEVPPNQHVVRKFFYLLEPRRACGTCPRDYVEINGEKYCGE 412
 QY 241 RSQFVVTNSNKTIVRHSQSYTDGTGFLAEVLSYDSDPCPQFTQRTQRCIRKELRCD 300
 DB 413 RSQFVVTNSNKTIVRHSQSYTDGTGFLAEVLSYDSDPCPQFTQRTQRCIRKELRCD 472
 QY 301 GMADCTHSDLELNCSCDAGHOFCKNFKCPFLFVVCDSVNDGDNSEDGSCSCPAQTFRC 360
 DB 473 GMADCTHSDLELNCSCDAGHOFCKNFKCPFLFVVCDSVNDGDNSEDGSCSCPAQTFRC 532
 QY 361 SNGKCLSKSQQCNKGKDCGDSDBASCPRKXNVVYCTKHYRCINGLCLSKGNPECDEKED 420
 DB 533 SNGKCLSKSQQCNKGKDCGDSDBASCPRKXNVVYCTKHYRCINGLCLSKGNPECDEKED 592
 QY 421 CSQGSDEKDCDCLGRSTTRQARVVGCTDADGEMPMQVSLHALGQGHICGASLISPMWLV 480
 DB 593 CSQGSDEKDCDCLGRSTTRQARVVGCTDADGEMPMQVSLHALGQGHICGASLISPMWLV 652
 QY 481 SAACHYIDDRGFRYSDDPTQWTAFLGLHDQSORSAQVQERLKRILISHPFNDFTFYDI 540
 DB 653 SAACHYIDDRGFRYSDDPTQWTAFLGLHDQSORSAQVQERLKRILISHPFNDFTFYDI 712
 QY 541 ALLELEKPAEYSSWVRPICLPDASHVFPAGKAIWVTGMGHTQYGGTALILQKEIRVIN 600
 DB 713 ALLELEKPAEYSSWVRPICLPDASHVFPAGKAIWVTGMGHTQYGGTALILQKEIRVIN 772
 QY 601 QTTCEBNLLPQOITPRMNCVGFSLSGVDSCGDSGGPLSSVEADGRIFQAGVWSMGDCAQ 660
 DB 773 QTTCEBNLLPQOITPRMNCVGFSLSGVDSCGDSGGPLSSVEADGRIFQAGVWSMGDCAQ 832
 QY 661 RNKPGVYTRLPFRDWMKENTGV 683
 DB 833 RNKPGVYTRLPFRDWMKENTGV 855

RESULT 21
 ID AAO30146
 AAO30146 standard; protein: 855 AA.
 XX
 AC AAO30146;
 DT 03-SEP-2003 (first entry)
 XX
 DE Human membrane-type serine protease MTSP1 protein.
 XX
 KM Serine protease 17; CYSPI7; tumour; cancer; antinease therapy; prostate;
 KM breast; cervix; lung; ovary; colon; gene therapy; human; enzyme; MTSP1;
 KM membrane-type serine protease; matrixase.

Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;			
QY	1	MAERVVMLPRARSLKSFVVTSVAFPTDSKTVQRTODNSCSFGIARGLAEIEMRFTTPG	60
DB	173	MAERVVMLPRARSLKSFVVTSVAFPTDSKTVQRTODNSCSFGIARGLAEIEMRFTTPG	232
QY	61	FPDSPPYAHARCOMALRGDADSVSLTFRSFDLASCDERSGDLVTYVNTLSPMEPHALVQ	120
DB	233	FPDSPPYAHARCOMALRGDADSVSLTFRSFDLASCDERSGDLVTYVNTLSPMEPHALVQ	292
QY	121	LCGTPPSYNLTFRHSQNVLLITLITNTERRHGFPEATFQLPMSSCGRLRAQGTEN	180
DB	293	LCGTPPSYNLTFRHSQNVLLITLITNTERRHGFPEATFQLPMSSCGRLRAQGTEN	352
QY	181	SPYYPGHYPNIDCTWNIIEVNNQHVYKRFKFLLEPRACGCPDYVEINEKTCGE	240
DB	353	SPYYPGHYPNIDCTWNIIEVNNQHVYKRFKFLLEPGVAGCPDYVEINEKTCGE	412
QY	241	RSQFVTSNSNKTIVRFHSDSYTDGFLAEYLSYSSDPCPGQFTCRTRCIRKEIRCD	300
DB	413	RSQFVTSNSNKTIVRFHSDSYTDGFLAEYLSYSSDPCPGQFTCRTRCIRKEIRCD	472
QY	301	GMADCTDHSDELNCSCDAGHQTCKNFKCKPLFWVCDVNDGDNDSDEQSCCPAQTFRC	360
DB	473	GMADCTDHSDELNCSCDAGHQTCKNFKCKPLFWVCDVNDGDNDSDEQSCCPAQTFRC	532
QY	361	SNKCLSKSQOCCNGKDCGDSDEASCPKXNVVTCIKTRKCLNGCLSKNPECCDGED	420
DB	533	SNKCLSKSQOCCNGKDCGDSDEASCPKXNVVTCIKTRKCLNGCLSKNPECCDGED	592
QY	421	CSDSDEKDCDCGRLSTFRQARVVGTDADGEMPMQVSLHALGQGHICGASLISPMVLV	480
DB	593	CSDSDEKDCDCGRLSTFRQARVVGTDADGEMPMQVSLHALGQGHICGASLISPMVLV	652
QY	481	SAACHYIDDRGFYSPTQWTAFGLHDQSORAPGVQERLKR1ISHPFNDFTFDYDI	540
DB	653	SAACHYIDDRGFYSPTQWTAFGLHDQSORAPGVQERLKR1ISHPFNDFTFDYDI	712
QY	541	ALLELEKPAEYSSMWRIICLPDASHVPAGKAIWVGWGHYQYGGTALILQKEIRVIN	600
DB	713	ALLELEKPAEYSSMWRIICLPDASHVPAGKAIWVGWGHYQYGGTALILQKEIRVIN	772
QY	601	QTTCEMLPQOITPRMNCVGLSGVDSQCGDSGGLSVYADGRIFQAGVSWGDCQAQ	660
DB	773	QTTCEMLPQOITPRMNCVGLSGVDSQCGDSGGLSVYADGRIFQAGVSWGDCQAQ	832
QY	661	RNKGVTYTRLPFLFDWIKENTGV	683
DB	833	RNKGVTYTRLPFLFDWIKENTGV	855

RESULT 23

AAE29791 ID AAE29791 standard; protein: 855 AA.

XX AAE29791:

XX 24-FEB-2003 (first entry)

XX Human membrane-type serine protease, MTSPL.

XX Human; serine protease 14; CUSP14; cancer; malignancy; breast; colon;

XX gene therapy; cytosolic; membrane-type serine protease; enzyme; MTSPL.

XX Homo sapiens.

XX OS

XX PN WO200277263-A2.

XX PD 03-OCT-2002.

XX XX

XX 20-MAR-2002; 2002WO-US009039.

XX PR 22-MAR-2001; 2001US-0278166P.

XX	(CORV-) CORVAS INT INC.
PA	
XX	Madison EL, Yeh J;
PI	
XX	WPI; 2003-018938/01.
DR	N-PSDB; AAD47180.
XX	New purified CUSP14 polypeptide and encoding nucleic acid molecule,
PT	useful for diagnosing, preventing and/or treating disorders, such as
PT	cancers and malignancies of the breast, cervix, prostate, lung, ovary or
XX	colon.
PS	Disclosure; Page 171-173; 185pp; English.
XX	
CC	The invention relates to transmembrane serine protease 14 (CUSP14), its
CC	nucleic acid sequence and the method based on them. The methods and
CC	compositions of the invention are useful for diagnosing, preventing
CC	and/or treating conditions associated with the aberrant expression or
CC	activity of the CUSP14 polypeptide, such as cancers and malignancies of
CC	the breast, cervix, prostate, lung, ovary or colon. The methods are also
CC	useful for identifying compounds that will modulate the protease activity
CC	of CUSP14 polypeptide, and monitoring tumour progression and/or
CC	therapeutic effectiveness. CUSP14 DNA used in gene therapy. The present
CC	sequence is human membrane-type serine protease, MTSPL
XX	
SO	Sequence 855 AA;
	Query Match 99.1%; Score 3754; DB 6; Length 855;
	Best Local Similarity 99.3%; Pred. No. 1.2e-240;
	Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY	1 MAERVVMLPPRARSLKSFVVTSVAFPTDSKTVQRTODNSCSFGLHARGVLEMFITPG 60
DB	173 MAERVVMLPPRARSLKSFVVTSVAFPTDSKTVQRTODNSCSFGLHARGVLEMFITPG 232
QY	61 PPDSPPYAHARCOMALRGDADSVSLTFRSFDLASCDERSGDLVTYVNTLSMMEHALVQ 120
DB	233 PPDSPPYAHARCOMALRGDADSVSLTFRSFDLASCDERSGDLVTYVNTLSMMEHALVQ 292
QY	121 LCGTPPSYNLTFRHSQNVLLITLITNTERRHGPGEATFQLPMSSCGRLRAQGTEN 180
DB	293 LCGTPPSYNLTFRHSQNVLLITLITNTERRHGPGEATFQLPMSSCGRLRAQGTEN 352
QY	181 SPYYPGHYPNIDCTWNIIEVNNQHVYKRFKFLLEPRACGTCPKDYVEINEKTCGE 240
DB	353 SPYYPGHYPNIDCTWNIIEVNNQHVYKRFKFLLEPGVAGTCPKDYVEINEKTCGE 412
QY	241 RSQFVTSNSNKTIVRFHSDSYTDGFLAEYLSYSSDPCPGQFTCRTRCIRKEIRCD 300
DB	413 RSQFVTSNSNKTIVRFHSDSYTDGFLAEYLSYSSDPCPGQFTCRTRCIRKEIRCD 472
QY	301 GMADCTDHSDELNCSCDAGHQTCKNFKCKPLFWVCDVNDGDNDSDEQSCCPAQTRFC 360
DB	473 GMADCTDHSDELNCSCDAGHQTCKNFKCKPLFWVCDVNDGDNDSDEQSCCPAQTRFC 532
QY	361 SNGKCLSKSQOCCNGKDCGDSDEASCPKXNVVTCIKTRYCLNGCLSKGNPECCDGED 420
DB	533 SNGKCLSKSQOCCNGKDCGDSDEASCPKXNVVTCIKTRYCLNGCLSKGNPECCDGED 592
QY	421 CSDGSDKDCDCGLRSFTFRQARVVGGTDADGEMPMQVSLHALGQGHICGASLISPMVLV 480
DB	593 CSDGSDKDCDCGLRSFTFRQARVVGGTDADGEMPMQVSLHALGQGHICGASLISPMVLV 652
QY	481 SAACHYIDDRGFYSPTQWTAFGLHDQSORAPGVQERLKLIISSHPFNDFTFDYDI 540
DB	653 SAACHYIDDRGFYSPTQWTAFGLHDQSORAPGVQERLKLIISSHPFNDFTFDYDI 712
QY	541 ALLELEKPAEYSSMWRIICLPDASHVPAGKAIWVGWGHYQYGGTALLILOKEIRIVIN 600
DB	713 ALLELEKPAEYSSMWRIICLPDASHVPAGKAIWVGWGHYQYGGTALLILOKEIRIVIN 772
QY	601 QTTCEMLPQOITPRMNCVGLSGGVDSQCGDSGGLSVYADGRIFQAGVWSWGDCQAQ 660

Db 773 OTTCENLIPQOITPRMWCVFLSGVDSCQDGSGLSVLEADGRIFQAGVSWGDCAQ 832
Qy 661 RNRKGVYTRLPFLRDMWKENTGV 683
Db 833 RNRKGVYTRLPFLRDMWKENTGV 855

RESULT 24

ABP72376
ID ABP72376 standard; protein; 855 AA.

AC ABP72376;

DT 13-MAY-2003 (first entry)

DE Transmembrane serine protease 1 (MTSP1).

XX Human; transmembrane serine protease 1; MTSP1; MTSP20; enzyme;
KM cytosolic; dermatological; cardiant; vulnerary; ophthalmological;
KM gene therapy.

OS Homo sapiens.

PN WO2003004681-A2.

XX 16-JAN-2003.

PF 03-JUL-2002; 2002WO-US021208.

XX 03-JUL-2001; 2001US-0302939P.

XX (CORV-) CORVAS INT INC.

PI Madison EL, Ong EO;

XX WPI; 2003-239207/23.

DR N-PSDB; ABZ58500.

PT New type-II membrane-type serine protease 20 polypeptides, useful for
preparing a medicament for diagnosing, treating or preventing cancer,
PT dermatological disorders, aberrant wound repairs or crest syndromes.

PS Disclosure; Page 198-200; 216pp; English.

XX The present sequence is the protein sequence of human type II
XX transmembrane serine protease 1 (MTSP1). The invention relates to novel
XX human type II transmembrane serine protein 20 (MTSP20) (see ABP72374) and
XX nucleic acids encoding it (see ABZ58499). Also claimed are methods of
XX inhibiting tumour initiation, growth or progression by inhibiting MTSP20
XX activity, and of treating or preventing a disease or disorder associated
XX with undesired and/or uncontrolled angiogenesis or neovascularisation,
XX especially undesired angiogenesis associated with solid neoplasms,
XX vascular malformations and cardiovascular disorders, chronic inflammatory
XX diseases, aberrant wound repairs, circulatory disorders, crest syndromes,
XX dermatological disorders and ocular disorders using an inhibitor of MSP20
XX
XX Sequence 855 AA;

Query Match 99.1%; Score 3754; DB 6; Length 855;
Best Local Similarity 99.3%; Pred. No. 1.2e-240;
Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 MAERVVMLPPRARSLSFVVTSVVAPPTDSTKYQRTQDNSSCFGLHARGVELMRFTTPG 60
Db 173 MAERVVMLPPRARSLSFVVTSVVAPPTDSTKYQRTQDNSSCFGLHARGVELMRFTTPG 232
Qy 61 FPDSPYPAHARCGWALRGDADSVLSTFRSFDLASCDERSGDLVTYNTLSPMEPHALVQ 120
Db 233 FPDSPYPAHARCGWALRGDADSVLSTFRSFDLASCDERSGDLVTYNTLSPMEPHALVQ 292
Qy 121 LCGTYPSTVNLTFHSSQNVLLITLINTERRHGFEEATFFQLPRMSSCGRLKRAQGTEN 180

Db 293 LCGTYPSTVNLTFHSSQNVLLITLINTERRHGFEEATFFQLPRMSSCGRLKRAQGTEN 352
Qy 181 SPYYPGHYPNIDCTNIEVPNNOHKVAEFKFFYLLEPRRACGTCPODYVEINGEKXCGE 240
Db 353 SPYYPGHYPNIDCTNIEVPNNOHKVAEFKFFYLLEPRRACGTCPODYVEINGEKXCGE 412
Qy 241 RSGFVVTSSNKTIVAFHSDQSYDTGFLAEYLSDSDPCPOQFTCRTRCIRKELRCD 300
Db 413 RSGFVVTSSNKTIVAFHSDQSYDTGFLAEYLSDSDPCPOQFTCRTRCIRKELRCD 472
Qy 301 GMADCTDHSDELNCSDDAGHOFCKRKFKCKPLFWVCDSVNDGDNSDEGSCCPAOTFRC 360
Db 473 GMADCTDHSDELNCSDDAGHOFCKRKFKCKPLFWVCDSVNDGDNSDEGSCCPAOTFRC 532
Qy 361 SNGKCLSKSQQCGKNGKDCDGSDEASCPRKYNVYTCFKHYRCNLGCLSKGNPECDKED 420
Db 533 SNGKCLSKSQQCGKNGKDCDGSDEASCPRKYNVYTCFKHYRCNLGCLSKGNPECDKED 592
Qy 421 CSDGDEKDCDGLRSFTROARVVGCTDADDEGEWPMQVSLHALGQGHICGASLISPMWLV 480
Db 593 CSDGDEKDCDGLRSFTROARVVGCTDADDEGEWPMQVSLHALGQGHICGASLISPMWLV 652
Qy 481 SAHNCYIDRGRYSDPTQWTAFLGLHDOSASAPGVQERRLKRIISHPFENDFTFDYDI 540
Db 653 SAHNCYIDRGRYSDPTQWTAFLGLHDOSASAPGVQERRLKRIISHPFENDFTFDYDI 712
Qy 541 ALLELEKPAEYSSMWVPICLPDASHVFPAGKAIWYTGWHTQYCGTGALILQKGEIRVIN 600
Db 713 ALLELEKPAEYSSMWVPICLPDASHVFPAGKAIWYTGWHTQYCGTGALILQKGEIRVIN 772
Qy 601 OTTCENLIPQOITPRMWCVFLSGVDSCQDGSGLSVLEADGRIFQAGVSWGDCAQ 660
Db 773 OTTCENLIPQOITPRMWCVFLSGVDSCQDGSGLSVLEADGRIFQAGVSWGDCAQ 832
Qy 661 RNRKGVYTRLPFLRDMWKENTGV 683
Db 833 RNRKGVYTRLPFLRDMWKENTGV 855

RESULT 25

ADB97551
ID ADB97551 standard; protein; 855 AA.

XX ADB97551;

AC 04-DEC-2003 (first entry)

DT Human MTSP1, SEQ ID NO:2.

XX Human; type I transmembrane serine protease 1; MTSP1; matrixpease;
KM serine protease; protease cleavage activation; diagnostic marker;
KM neoplastic disorder; cancer; breast; cervix; prostate; lung; ovary;
KM cytosolic; gene therapy; drug screening; tumour progression; monitoring;
KM enzyme.

XX Homo sapiens.

PN WO2003031585-A2.

XX 17-APR-2003.

PF 08-OCT-2002; 2002WO-US032417.

PR 09-OCT-2001; 2001US-0328530P.

PA (CORV-) CORVAS INT INC.

PI Madison EL, Yeh J;

XX WPI; 2003-393442/37.

DR N-PSDB; ADB97550.

PT New purified single- or two-chain polypeptide, useful for diagnosing,

PT preventing or treating cancer (e.g. colon cancer), comprises a protease
 PT domain of a type-1 membrane-type serine protease 25 or its catalytically
 PT active portion.
 XX
 XX Disclosure; Page 179-181; 97pp; English.
 XX
 CC The invention relates to human type I transmembrane serine protease 25
 CC (MSP25; AD897565) and polypeptides derived from it (e.g., AD897555). The
 CC MSP25 gene is located on chromosome 12. MSP25 is a serine protease
 CC which is activated by protease cleavage to yield a two-chain protease
 CC comprising an A chain and a B chain linked by a disulphide bond. MSP25
 CC is expressed or is active in tumour cells, and can therefore be used as a
 CC diagnostic marker for certain cancers. The invention also encompasses
 CC nucleic acids encoding an MSP25 polypeptide (AD897564, AD897571).
 CC AD897554); nucleic acid vectors and host cells comprising an MSP25
 CC polynucleotide; a MSP25 knockout animal; and an antibody specific for
 CC either the single chain (zymogen) or two-chain (activated) form of
 CC MSP25. MSP25 polypeptides are useful in diagnosing, preventing or
 CC treating neoplastic diseases, such as cancer of the breast, cervix,
 CC prostate, lung, ovary or colon. MSP25 polypeptides may also be used for
 CC identifying compounds that modulate the protease activity of the
 CC polypeptide and for monitoring tumour progression and/or therapeutic
 CC effectiveness. The present sequence represents the related protein, MTSpl
 CC (also known as matrilysin).
 CC
 XX
 SQ Sequence 855 AA;

Query Match 99.1%; Score 3754; DB 7; Length 855;
 Best Local Similarity 99.3%; Pred. No. 1.2e-240;
 Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAERVVMLEPRARSLKSFVTVSVAPPTDSKTQVORTDQNSCSFGHARGVELMRFTTPG 60
 DB 173 MAERVVMLEPRARSLKSFVTVSVAPPTDSKTQVORTDQNSCSFGHARGVELMRFTTPG 232
 QY 61 FPDSPYPAHARCMQALRGDADSVLSTFRSFDLASCBERSGDVLTVVNTLSPMEPHALVQ 120
 DB 233 FPDSPYPAHARCMQALRGDADSVLSTFRSFDLASCBERSGDVLTVVNTLSPMEPHALVQ 292
 QY 121 LCGTTPSPSYNLTFHSSQNVLLITLITNTERRHPEFATFQGLPRMSSCGRLRRAQGTFN 180
 DB 293 LCGTTPSPSYNLTFHSSQNVLLITLITNTERRHPEFATFQGLPRMSSCGRLRRAQGTFN 352
 QY 181 SPYRGHYPPNIDCTNMITEVNNQNVYRKFETFLSPRACGTCPPDYVEINEKXCGE 240
 DB 353 SPYRGHYPPNIDCTNMITEVNNQNVYRKFETFLSPRACGTCPPDYVEINEKXCGE 412
 QY 241 RSQFVTVSNNSKLTFRFHSDDSYDTGTFLAEVLSYDSSDPCPGQFTCRGTGCIKEIRCD 300
 DB 413 RSQFVTVSNNSKLTFRFHSDDSYDTGTFLAEVLSYDSSDPCPGQFTCRGTGCIKEIRCD 472
 QY 301 GMADCTDHSDELNSCDAGHQFTCKNPKCKPLFWVCDNVNDGDNDEGSCCPAQTFRC 360
 DB 473 GMADCTDHSDELNSCDAGHQFTCKNPKCKPLFWVCDNVNDGDNDEGSCCPAQTFRC 532
 QY 361 SNGKCLSKSQOQCNCKDCCGSDASCPKVVVCTGTRKCLNGCLLSKKNPCDCKED 420
 DB 533 SNGKCLSKSQOQCNCKDCCGSDASCPKVVVCTGTRKCLNGCLLSKKNPCDCKED 592
 QY 421 CSDSDSKDCCDGGARSTFRQARVVGGTDADGEMPMOVSJHALGQGHICGASLSIPMWLV 480
 DB 593 CSDSDSKDCCDGGARSTFRQARVVGGTDADGEMPMOVSJHALGQGHICGASLSIPMWLV 652
 QY 481 SAACHYIDDRGFYSDEPTQWTAFLGLHDQGRSAPGVQERRLKRIISHPEFNDFTPDYDI 540
 DB 653 SAACHYIDDRGFYSDEPTQWTAFLGLHDQGRSAPGVQERRLKRIISHPEFNDFTPDYDI 712
 QY 541 ALLELEKPAEYSSNVRLCLPDASHVBPAGKAIWVTMGHTQVCGTALTIIQKEIRVIN 600
 DB 713 ALLELEKPAEYSSNVRLCLPDASHVBPAGKAIWVTMGHTQVCGTALTIIQKEIRVIN 772
 QY 601 QTTCENTLLPOQITPRMMKCVGFLSGVDSCOGDSGGLSVSEADGRIFQAGVSVSGDCAQ 660

DB 773 QTTCENTLLPOQITPRMMKCVGFLSGVDSCOGDSGGLSVSEADGRIFQAGVSVSGDCAQ 832
 QY 661 RNKRGVYTRPLPRDWTAKENTGV 683
 DB 833 RNKRGVYTRPLPRDWTAKENTGV 855

RESULT 26

AD110371
 ID AD110371 standard; protein; 855 AA.

XX
 AC AD110371;
 XX

DT 22-Apr-2004 (first entry)
 XX

DE Human cell surface protease #1.
 XX

XX therapeutic agent; plasmin; protease specific antigen; PSA;
 KM cell-surface protease-associated disease; cancer; ocular disease;

KM cardiovascular disease; chronic inflammatory disease; wound;
 KM circulatory disorder; dermatological disorder; rheumatoid arthritis;

KM psoriasis; diabetic retinopathy; pterygium;
 KM excimer laser surgery scarring; glaucoma filtering surgery scarring;

KM macular degeneration; cleft syndrome; solid neoplasm; vascular tumour;
 KM melanoma; Kaposi's sarcoma; human; cell surface protease.

XX Homo sapiens.

OS
 XX W0200295007-A2.
 XX

PD 28-NOV-2002.
 XX

PF 23-MAY-2002; 2002MO-USO16819.
 XX

PR 23-MAY-2001; 2001US-0293267P.
 XX

PA (CORV-) CORVAS INT INC.
 XX

PI Madison EL, Sempke JE, Vlausk GP, Kemp SJ, Komandla M, Siev DV;
 XX

XX WPI; 2003-221280/21.
 DR

DR N-PSDB; AD110370.
 XX

PT Novel conjugate useful for treating cell-surface protease-associated
 PT disease, comprises a therapeutic agent and a peptidic or nucleic acid

PT substrate linked to it optionally by a peptidic linker.
 XX

PS Claim 9; SEQ ID NO 2; 581pp; English.
 XX

CC The invention comprises a conjugate that consists of a therapeutic agent
 CC and a peptide substrate (optionally linked via linker). The peptide

CC substrate is proteolytically cleaved by a cell surface protease pr a
 CC soluble, released or shed form of it, to liberate the therapeutic agent,

CC the conjugate of the invention is not substantially cleaved by plasmin or
 CC protease specific antigen (PSA). The conjugate of the invention is useful

CC for treating a cell-surface protease-associated disease such as: cancer;
 CC ocular diseases, cardiovascular diseases, chronic inflammatory diseases,

CC wounds, circulatory disorders, dermatological disorders, rheumatoid
 CC arthritis, psoriasis, diabetic retinopathies, recurrence of pterygium,

CC scarring from excimer laser surgery, scarring from glaucoma filtering
 CC surgery, macular degeneration, cleft syndrome, solid neoplasms, vascular

CC tumours, melanoma and Kaposi's sarcoma. The present amino acid sequence
 CC represents a human cell surface protease.
 XX

SQ Sequence 855 AA;

Query Match 99.1%; Score 3754; DB 7; Length 855;
 Best Local Similarity 99.3%; Pred. No. 1.2e-240;
 Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAERVVMLEPRARSLKSFVTVSVAPPTDSKTQVORTDQNSCSFGHARGVELMRFTTPG 60
 DB 173 MAERVVMLEPRARSLKSFVTVSVAPPTDSKTQVORTDQNSCSFGHARGVELMRFTTPG 232

```

QY 61 FPDSPYPAHARCMQWALRGDADSVLSLTFRSPDLASCDERGSIDLVTYNTLSPMEPHALVQ 120
D 233 FPDSPYPAHARCMQWALRGDADSVLSLTFRSPDLASCDERGSIDLVTYNTLSPMEPHALVQ 292
QY 121 LCGTTPPSYNULTFRHSQNVLLITLITNTERRHGFPAATFQQLPRMSSCGRLRAQGTEN 180
D 293 LCGTTPPSYNULTFRHSQNVLLITLITNTERRHGFPAATFQQLPRMSSCGRLRAQGTEN 352
QY 181 SPYYPGHYPNIDCTWNIIEVNNQHVAFKFFFLLEPRRACGCPDQVYEINKEKTCGE 240
D 353 SPYYPGHYPNIDCTWNIIEVNNQHVAFKFFFLLEPRRACGCPDQVYEINKEKTCGE 412
QY 241 RSQFVVTSSNKKITVRFHSDQSYTDGTGFLAEYLSYDSSDPCEQFCTGRCIRKELRCD 300
D 413 RSQFVVTSSNKKITVRFHSDQSYTDGTGFLAEYLSYDSSDPCEQFCTGRCIRKELRCD 472
QY 301 GMAUCTHSDBELNCSQAGHQFTCKNFKCPFLFWVCDNVNDCGNSDEQSCGPAQTFRG 360
D 473 GMAUCTHSDBELNCSQAGHQFTCKNFKCPFLFWVCDNVNDCGNSDEQSCGPAQTFRG 532
QY 361 SNGKCLSKSQOQCKGKDCGDSDEASCPKYNVVTCTHTRRCNLGLCLSKGNPECDKED 420
D 533 SNGKCLSKSQOQCKGKDCGDSDEASCPKYNVVTCTHTRRCNLGLCLSKGNPECDKED 592
QY 421 CSDGSEKDCDCGLRFTROARVVGTDADGEMPMQVSLHALGQGHICGASLISPMNLV 480
D 593 CSDGSEKDCDCGLRFTROARVVGTDADGEMPMQVSLHALGQGHICGASLISPMNLV 652
QY 481 SAHACYIDDRGFRYSPTQMTAFGLHDSORSAPGVQERRLRKRIISHPFNDFTFYDI 540
D 653 SAHACYIDDRGFRYSPTQMTAFGLHDSORSAPGVQERRLRKRIISHPFNDFTFYDI 712
QY 541 ALLELEKPAEYSSWVRPCLPDASHVFPAGKAIWVTGMGHTQYGGTGALLIQQEIRVIN 600
D 713 ALLELEKPAEYSSWVRPCLPDASHVFPAGKAIWVTGMGHTQYGGTGALLIQQEIRVIN 772
QY 601 QTTTCENILPQOITPRMMCVGFLSGVDVSCQDGSGLPSSVEADGRIFQAGVNSMGDCAQ 660
D 773 QTTTCENILPQOITPRMMCVGFLSGVDVSCQDGSGLPSSVEADGRIFQAGVNSMGDCAQ 832
QY 661 RNKPGVYTRLP.LPRDWTIKENTGV 683
D 833 RNKPGVYTRLP.LPRDWTIKENTGV 855

```

RESULT 27

ADG65326

ADG65326 standard; protein; 855 AA.

AC ADG65326;

11-MAR-2004 (first entry)

Human MTSPI.

human, type II membrane-type serine protease 12; MTSPI2; chromosome 19;

protease domain; MTSPI2-PD1; MTSPI2-PD2; MTSPI2-PD3; lung; oesophagal;

prostate; colon; ovary; cervix; breast; pancreas; cancer; tumour;

neoplastic condition.

Homo sapiens.

Key Location/Qualifiers

Domain 615..855

21-MAY-2003; 2003MO-US016181.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

PR 21-MAY-2002; 2002US-0382851P.

(DEND-) DENDREON SAN DIEGO LLC.

Medison EL, Ong EO;

WPI; 2004-062325/06.

New type II membrane-type serine protease 12 proteins and nucleic acids,

useful in diagnostics, particularly for diagnosing lung, esophagal,

prostate, colon, ovary, cervix, breast and pancreas cancers.

Disclosure; SEQ ID NO 2; 207bp; English.

This sequence represents a human type II membrane-type serine protease 1

(MTSP1). The MTSPI protein sequence was used in the isolation of the

coding sequence of MTSPI2, which has been isolated to chromosome 19.

MTSP12 does not include the sequence of amino acids Arg-Lys-His-Pro-

Arg-Pro-Ala ADG65347. The MTSPI2 coding domain of MTSPI to search the human

HTGS database which produced three serine proteases, MTSPI2-PD1, MTSPI2-

PD2 and MTSPI2-PD3. EST's were identified which matched portions of

MTSP12-PD1, MTSPI2-PD2 and MTSPI2-PD3. MTSPI2 polypeptides, proteins and

nucleic acids are useful in diagnostics, particularly for diagnosing

lung, oesophagal, prostate, colon, ovary, cervix, breast and pancreas

cancers. These are useful in immunoassays to detect, prognosis, diagnose,

or monitor various conditions, diseases, and disorders affecting MTSPI2

polypeptide expression, or monitor their treatment. Modulators of MTSPI2

are useful for treating cancer, tumour and other neoplastic conditions.

Sequence 855 AA:

Query Match 99.1%; Score 3754; DB 8; Length 855;

Best Local Similarity 99.3%; Pred. No. 1,2e-240;

Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

QY 1 MAERVYMLPPRARSLKSTVYTSVVAFPPTDSKVQRTQNSCGFGLHARGVELMRFTTGG 60
D 173 MAERVYMLPPRARSLKSTVYTSVVAFPPTDSKVQRTQNSCGFGLHARGVELMRFTTGG 232
QY 61 FPDSPYPAHARCMQWALRGDADSVLSLTFRSPDLASCDERGSIDLVTYNTLSPMEPHALVQ 120
D 233 FPDSPYPAHARCMQWALRGDADSVLSLTFRSPDLASCDERGSIDLVTYNTLSPMEPHALVQ 292
QY 121 LCGTTPPSYNULTFRHSQNVLLITLITNTERRHGFPAATFQQLPRMSSCGRLRAQGTEN 180
D 293 LCGTTPPSYNULTFRHSQNVLLITLITNTERRHGFPAATFQQLPRMSSCGRLRAQGTEN 352
QY 181 SPYYPGHYPNIDCTWNIIEVNNQHVAFKFFFLLEPRRACGCPDQVYEINKEKTCGE 240
D 353 SPYYPGHYPNIDCTWNIIEVNNQHVAFKFFFLLEPRRACGCPDQVYEINKEKTCGE 412
QY 241 RSQFVVTSSNKKITVRFHSDQSYTDGTGFLAEYLSYDSSDPCEQFCTGRCIRKELRCD 300
D 413 RSQFVVTSSNKKITVRFHSDQSYTDGTGFLAEYLSYDSSDPCEQFCTGRCIRKELRCD 472
QY 301 GMAUCTHSDBELNCSQAGHQFTCKNFKCPFLFWVCDNVNDCGNSDEQSCGPAQTFRG 360
D 473 GMAUCTHSDBELNCSQAGHQFTCKNFKCPFLFWVCDNVNDCGNSDEQSCGPAQTFRG 532
QY 361 SNGKCLSKSQOQCKGKDCGDSDEASCPKYNVVTCTHTRRCNLGLCLSKGNPECDKED 420
D 533 SNGKCLSKSQOQCKGKDCGDSDEASCPKYNVVTCTHTRRCNLGLCLSKGNPECDKED 592
QY 421 CSDGSEKDCDCGLRFTROARVVGTDADGEMPMQVSLHALGQGHICGASLISPMNLV 480
D 593 CSDGSEKDCDCGLRFTROARVVGTDADGEMPMQVSLHALGQGHICGASLISPMNLV 652
QY 481 SAHACYIDDRGFRYSPTQMTAFGLHDSORSAPGVQERRLRKRIISHPFNDFTFYDI 540
D 653 SAHACYIDDRGFRYSPTQMTAFGLHDSORSAPGVQERRLRKRIISHPFNDFTFYDI 712
QY 541 ALLELEKPAEYSSWVRPCLPDASHVFPAGKAIWVTGMGHTQYGGTGALLIQQEIRVIN 600

```

```
Db 713 ALLELEKPAEYSSWVRPCLPDAASHVPAGKAIWVTGHTGYCGTALLIQKEIRVIN 772
Qy 601 QTTCENLTPQOITPRMNCVGFSLSGVDS CGDSCGSPLSVEADGRITFOAGVWSMGDCAQ 660
Db 773 QTTCENLTPQOITPRMNCVGFSLSGVDS CGDSCGSPLSVEADGRITFOAGVWSMGDCAQ 832
Qy 661 RNKGVYTRLP LPRDWIKENTGV 683
Db 833 RNKGVYTRLP LPRDWIKENTGV 855

RESULT 28
ADJ28861
ID ADJ28861 standard; protein; 855 AA.
XX
AC ADJ28861;
XX
DT 22-APR-2004 (first entry)
XX
DE Human matricinase ( MTSP1) serine protease.
XX
KM Human; matricinase; MTSP1; CVPSP16; cytostatic; gene therapy; tumour;
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Domain 615..855
FT /note= "Protease domain"
XX
PN NO2004005471-A2.
XX
PD 15-JAN-2004.
XX
PF 01-JUL-2003; 2003WO-US020959.
XX
PR 02-JUL-2002; 2002US-0394347P.
XX
PA (DEND-) DENDREON SAN DIEGO LLC.
XX
PI Madison EL, Ong EO, Yeh J;
XX
DR WPI; 2004-099379/10.
DR N-PSDB; ADJ28860.
XX
PT New serine protease 16 polypeptides, useful for diagnosing, preventing or
PT treating cancer (e.g. breast cancer) or for identifying compounds that
PT may be used for modulating the polypeptides.
XX
PS Disclosure; SEQ ID NO 2; 205BP; English.
XX
CC The present sequence is the protein sequence of human MTSP1 (matricinase),
CC a serine protease that is expressed in epithelial cancer and normal
CC tissue. In an example from the invention, the protein sequence of the
CC protease domain of MTSP1 was used to search a human genome database. A
CC protease was identified and designated serine protease 16 or CVPSP16.
CC CVPSP16 cDNA ADJ28863 was subsequently cloned from a human liver cDNA
CC library by PCR. CVPSP16 polypeptides exhibit protease activity as a single
CC chain or as a multi-chain form. Methods are provided for identifying
CC compounds that modulate the protease activity. CVPSP16 polypeptides also
CC serve as tumour markers.
XX
SQ Sequence 855 AA;
Qy Query Match 99.1%; Score 3754; DB 8; Length 855;
Best Local Similarity 99.3%; Pred. No. 1.2e-240;
Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Qy 1 MAERVVMLPFRASLSFVVTSVVAFPTDSKTVQRTDNCSCFGLHARGVELMRFTTPG 60
Db 173 MAERVVMLPFRASLSFVVTSVVAFPTDSKTVQRTDNCSCFGLHARGVELMRFTTPG 232
```

```
Qy 61 PPDSPPAHARCOMALRGDADSVLSLTFRSFDLASCDERSDLYVTYNTLSPMEPHALVQ 120
Db 233 PPDSPPAHARCOMALRGDADSVLSLTFRSFDLASCDERSDLYVTYNTLSPMEPHALVQ 292
Qy 121 LCGTTPPSYNULTFHSQGNVLLITLITNTERRHQFEATFQLPMSSCGRLRAQGTFN 180
Db 293 LCGTTPPSYNULTFHSQGNVLLITLITNTERRHQFEATFQLPMSSCGRLRAQGTFN 352
Qy 181 SPYTPGHYPPRIDCTWNIIEVPPNQHVVRKFPYLLPRAAGTCPPDYIENGKXCGE 240
Db 353 SPYTPGHYPPRIDCTWNIIEVPPNQHVVRKFPYLLPRAAGTCPPDYIENGKXCGE 412
Qy 241 RSOPVVTNSNKITTVRPHSDQSYTDPTGFLAEVLSYDSDPCPGQFTRGRICRKELRCD 300
Db 413 RSOPVVTNSNKITTVRPHSDQSYTDPTGFLAEVLSYDSDPCPGQFTRGRICRKELRCD 472
Qy 301 GMADCTDHSDELNCSCDAGHQFTCKNFKCP LFWVCDVNDGDNDSDEGSCSPAQTFRC 360
Db 473 GMADCTDHSDELNCSCDAGHQFTCKNFKCP LFWVCDVNDGDNDSDEGSCSPAQTFRC 532
Qy 361 SNGKCLSKSQOCCNGKDDCGSDBASCPKXNVVYTCYHTYRCINGLCLSKGNPECDKED 420
Db 533 SNGKCLSKSQOCCNGKDDCGSDBASCPKXNVVYTCYHTYRCINGLCLSKGNPECDKED 592
Qy 421 CSDSDEKDCDCGLRSPTROARVVGTDADGEMPMQVSLHALGQGHICASLISPMVLV 480
Db 593 CSDSDEKDCDCGLRSPTROARVVGTDADGEMPMQVSLHALGQGHICASLISPMVLV 652
Qy 481 SAACHYIDRGRFYSDEPTQWTAFLGLHDQSQSAPOVERRLKRIISHPFNDFTFYDI 540
Db 653 SAACHYIDRGRFYSDEPTQWTAFLGLHDQSQSAPOVERRLKRIISHPFNDFTFYDI 712
Qy 541 ALLELEKPAEYSSWVRPCLPDAASHVPAGKAIWVTGHTGYCGTALLIQKEIRVIN 600
Db 713 ALLELEKPAEYSSWVRPCLPDAASHVPAGKAIWVTGHTGYCGTALLIQKEIRVIN 772
Qy 601 QTTCENLTPQOITPRMNCVGFSLSGVDS CGDSCGSPLSVEADGRITFOAGVWSMGDCAQ 660
Db 773 QTTCENLTPQOITPRMNCVGFSLSGVDS CGDSCGSPLSVEADGRITFOAGVWSMGDCAQ 832
Qy 661 RNKGVYTRLP LPRDWIKENTGV 683
Db 833 RNKGVYTRLP LPRDWIKENTGV 855

RESULT 29
ADJ46895
ID ADJ46895 standard; protein; 855 AA.
XX
AC ADJ46895;
XX
DT 06-MAY-2004 (first entry)
XX
DE Human transmembrane serine protease (MTSP) polypeptide #1.
XX
CC Human; transmembrane serine protease; MTSP; cell surface protease;
CC plasmin; prostate specific antigen; PSA; proliferative disease;
CC cell-surface protease-associated disease; autoimmune disease;
CC inflammatory disease; infectious disease; endocrine disease; cancer;
CC ocular disorder; cardiovascular disorder; chronic inflammatory disease;
CC wound; circulatory disorder; dermatological disorder; restenosis;
CC rheumatoid arthritis; psoriasis; diabetic retinopathy;
CC laser surgery scarring; glaucoma filtering surgery scarring;
CC macular degeneration; CRST syndrome; bacterial infection; viral disease;
CC solid neoplasm; vascular tumour; lung; colon; prostate; melanoma;
CC Kaposi's sarcoma; enzyme.
XX
OS Homo sapiens.
XX
PN US2004001801-A1.
XX
PD 01-JAN-2004.
```

PF 23-MAY-2002; 2002US-00156214.
XX 23-MAY-2002; 2002US-00156214.
XX (CORV-) CORVAS INT INC.
PA
XX Madison EL, Semple JE, Vlausk GP, Kemp SJ, Komandla M, Slev DV;
XX WPI; 2004-190126/18.
DR N-PSDB; ADJ46894.
XX
XX Conjugate useful for treating e.g. cancer, cell-surface protease-
PT associated diseases, comprising a peptidic substrate or nucleic acid
PT substrate linked to a therapeutic agent through a linker.
PS
PS Claim 9; SEQ ID NO 2; 361pp; English.
XX
XX The invention relates to a conjugate comprising a therapeutic agent and a
CC peptidic substrate or nucleic acid substrate linked to the agent
CC optionally through a linker or peptidic linker, where the peptidic
CC substrate is proteolytically cleaved by a cell surface protease or a
CC soluble, released or shed form conjugate to liberate the agent and the
CC conjugate is not substantially cleaved by plasmin or prostate specific
CC antigen (PSA). The conjugate is useful for treating a disease, which
CC involves administering a conjugate to a subject, where the disease is
CC preferably a proliferative disease or a cell-surface protease-associated
CC disease. The diseases include autoimmune diseases, inflammatory diseases,
CC infectious diseases and endocrine diseases. The conjugate is useful for
CC treating a cell-surface protease-associated disease, which involves
CC administering a conjugate comprising an agent and a peptidic substrate to
CC a subject exhibiting symptoms of a cell-surface protease-associated
CC disorder, where the disease is selected from cancer, ocular disorders,
CC cardiovascular disorders, chronic inflammatory diseases, wounds,
CC circulatory disorders, dermatological disorders, restenosis, rheumatoid
CC arthritis, psoriasis, diabetic retinopathies, scarring from laser
CC surgery, scarring from glaucoma filtering surgery, macular degeneration,
CC CREST syndrome, bacterial infections, viral diseases, solid neoplasms and
CC melanoma and Kaposi's sarcoma. This sequence represents a transmembrane
CC serine protease (MTSP) polypeptide of the invention.
XX
SQ Sequence 855 AA;
Query Match 99.1%; Score 3754; DB 8; Length 855;
Best Local Similarity 99.3%; Pred. No. 1.2e-240;
Matches 678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 MAERVVMLEPPRARSLKSPVTVSVAPPTSKTVQRTQDNSCSFGLHARGVELMRFTTPG 60
DB 173 MAERVVMLPPRARSLKSPVTVSVAPPTSKTVQRTQDNSCSFGLHARGVELMRFTTPG 232
QY 61 FPDSPYPAHARCGWALRGDASVLSLTFRSFDLASCDERSDVLTVVNTLSPEPHALVQ 120
DB 233 FPDSPYPAHARCGWALRGDASVLSLTFRSFDLASCDERSDVLTVVNTLSPEPHALVQ 292
QY 121 LCGTYPSPYNLTFHSSQNVLLITLITNTERRHQFEATFQLPMSSCGRLRKAQGTEN 180
DB 293 LCGTYPSPYNLTFHSSQNVLLITLITNTERRHQFEATFQLPMSSCGRLRKAQGTEN 352
QY 181 SPVYRGHPYPRNIDCTWNIIEVNNQHVSRKFPFLLEPRACGTCPPDYVEINKEKCGE 240
DB 353 SPVYRGHPYPRNIDCTWNIIEVNNQHVSRKFPFLLEPRACGTCPPDYVEINKEKCGE 412
QY 241 RSQFVTVSNENKLTFRFHSQSYTDTFGLAEVLSYDSDDCPQGFRTGRCTRKELRCD 300
DB 413 RSQFVTVSNENKLTFRFHSQSYTDTFGLAEVLSYDSDDCPQGFRTGRCTRKELRCD 472
QY 301 GWADCTDHSDELNCSGDAHQFTCKNFKCKPLFWVCDVNDCCGNSDEGSCCEPAQTRFC 360
DB 473 GWADCTDHSDELNCSGDAHQFTCKNFKCKPLFWVCDVNDCCGNSDEGSCCEPAQTRFC 532
QY 361 SNGKCLSKSQOQCKDCCGSDGDEASCPKXNVVTCGHTTRCLNGLCLSKGNPECDCKED 420
XX

DB 533 SNGKCLSKSQOQCKDCCGSDGDEASCPKXNVVTCGHTTRCLNGLCLSKGNPECDCKED 592
QY 421 CSDGDEKDCDCGLRSFTRQARVVGCTDADDEGEMPVQVSLHALGQGHICGASLISPNMLV 480
DB 593 CSDGDEKDCDCGLRSFTRQARVVGCTDADDEGEMPVQVSLHALGQGHICGASLISPNMLV 652
QY 481 SAANCYIDDRGFRYSPTQMTAFGLHDSQSRSAPEVOERRLKRIISHPEFNDFTDYDI 540
DB 653 SAANCYIDDRGFRYSPTQMTAFGLHDSQSRSAPEVOERRLKRIISHPEFNDFTDYDI 712
QY 541 ALLELEKPAEYSWMPICLPDASHVFPAGKAIWVGWGHITQYGGGALIILOKEIRVIN 600
DB 713 ALLELEKPAEYSWMPICLPDASHVFPAGKAIWVGWGHITQYGGGALIILOKEIRVIN 772
QY 601 QTTCEMLPQOITPRMNCVGLSGVDSCQSDSGPGLSVEADGRIFOAGVSWBGCAQ 660
DB 773 QTTCEMLPQOITPRMNCVGLSGVDSCQSDSGPGLSVEADGRIFOAGVSWBGCAQ 832
QY 661 RNKPGYVTRLPFRDWIKENTGV 683
DB 833 RNKPGYVTRLPFRDWIKENTGV 855
RESULT 30
AAM25628
ID AAM25628 standard; protein; 851 AA.
XX
XX AAM25628;
AC
XX
XX 16-OCT-2001 (first entry)
DT
XX
XX Human protein sequence SEQ ID NO:1143.
DE
XX
XX Human; cancer; HIV infection; human immunodeficiency virus;
KW anti-inflammation; anti-rheumatic; antiarthritic; immunosuppressive;
KW antibacterial; endocrine; cardiac; central nervous system; viricide;
KW anti-HIV; fungicide; antitumor; cardiovascular; antineoplastic; anaemia;
KW antileishmaniasis; haemostatic; vulnery; antitumor; osteopathic; eczema;
KW dermatological; anti-allergic; antisthmatic; antidiabetic; cytostatic;
KW neuroprotective; antidepressant; neurotropic; antiparkinsonian; infection;
KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
KW cardiac dysfunction; neuropathology; cardiac amyloidosis; autoimmunity;
KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
KW thrombocytopenia; osteoporosis; severe combined immunodeficiency;
KW allergic rhinitis; diabetes; multiple sclerosis; depression;
KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
KW neurological disorder.
XX
XX Homo sapiens.
OS
XX
XX MO200153455-A2.
PN
XX
XX 26-JUL-2001.
PD
XX
XX 22-DEC-2000; 2000MO-US035017.
PF
XX
XX 23-DEC-1999; 99US-00471275.
PR 21-JAN-2000; 2000US-00488725.
PR 25-APR-2000; 2000US-00552317.
XX
XX (HYSB-) HYSBQ INC.
PA
XX
XX Tang YT, Liu C, Drmanac RT;
PI
XX
XX WPI; 2001-457603/49.
DR N-PSDB; AAM93569.
XX
XX Isolated human polynucleotides encoding polypeptides, useful for the
PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection.
XX
XX Claim 20; Page 237; 1217pp; English.
XX

This Page Blank (uspto)